

STIC-Biotech/ChemLib

49087

From: Mertz, Prema
Sent: Thursday, August 16, 2001 11:22 AM
To: STIC-Biotech/ChemLib
Subject: 09/521,195

Please search SEQ ID NO:1, 3, with protein databases.

Thanks
Prema Mertz, Ph.D.
Art Unit 1646
Crystal Mall 1, Room 10E-01
United States Patent & Trademark Office
(703) 308-4229

CRFE

*Edward Hart
Technical Info Specialist
STIC / Biotech
CM1 12C14 Tel: 305-9293*

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:55:18 ; Search time 20.69 Seconds
(without alignments)
548.346 Million cell updates/sec

Title: US-09-521-195-1
Perfect score: 2845
Sequence: 1 MRDYDEVIAFLGEGWPFQRL.....KKTRDSMETENPKVLITAF 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|------------------|
| 1 | 702 | 24.7 | 555 | 3 | US-08-501-572-3 |
| 2 | 702 | 24.7 | 555 | 3 | US-09-040-444-3 |
| 3 | 687.5 | 24.2 | 556 | 3 | US-08-501-572-1 |
| 4 | 687.5 | 24.2 | 556 | 3 | US-09-040-444-1 |
| 5 | 680 | 23.9 | 553 | 3 | US-08-501-572-2 |
| 6 | 680 | 23.9 | 553 | 3 | US-09-040-444-2 |
| 7 | 630 | 22.1 | 537 | 2 | US-08-647-397-2 |
| 8 | 234.5 | 8.2 | 494 | 2 | US-09-031-392-5 |
| 9 | 234.5 | 8.2 | 494 | 2 | US-09-299-549-5 |
| 10 | 222 | 7.8 | 492 | 2 | US-08-355-844-3 |
| 11 | 222 | 7.8 | 492 | 5 | PCT-US95-16126-3 |
| 12 | 216 | 7.6 | 493 | 2 | US-09-031-392-10 |
| 13 | 216 | 7.6 | 493 | 4 | US-09-299-549-10 |
| 14 | 213.5 | 7.5 | 524 | 2 | US-08-928-692-12 |
| 15 | 196 | 6.9 | 500 | 2 | US-09-031-392-7 |
| 16 | 196 | 6.9 | 500 | 4 | US-09-299-549-7 |
| 17 | 181 | 6.4 | 509 | 2 | US-09-031-392-6 |
| 18 | 181 | 6.4 | 509 | 4 | US-09-299-549-6 |
| 19 | 174.5 | 6.1 | 584 | 2 | US-08-928-692-13 |
| 20 | 172 | 6.0 | 383 | 2 | US-09-031-392-3 |
| 21 | 172 | 6.0 | 383 | 4 | US-09-299-549-3 |
| 22 | 169.5 | 6.0 | 488 | 2 | US-08-928-692-11 |
| 23 | 168.5 | 5.9 | 488 | 2 | US-08-928-692-10 |
| 24 | 162 | 5.7 | 563 | 2 | US-09-031-392-2 |
| 25 | 162 | 5.7 | 563 | 4 | US-09-299-549-2 |
| 26 | 158.5 | 5.6 | 109 | 2 | US-08-647-397-4 |
| 27 | 156.5 | 5.5 | 534 | 2 | US-09-031-392-4 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 28 | 156.5 | 5.5 | 534 | 4 | US-09-299-549-4 | Sequence 4, Appli |
| 29 | 110 | 3.9 | 473 | 1 | US-08-597-236-13 | Sequence 13, Appli |
| 30 | 110 | 3.9 | 473 | 1 | US-08-746-682A-13 | Sequence 13, Appli |
| 31 | 110 | 3.9 | 548 | 3 | US-08-903-139B-9 | Sequence 9, Appli |
| 32 | 109.5 | 3.8 | 834 | 2 | US-08-677-734A-9 | Sequence 9, Appli |
| 33 | 109.5 | 3.8 | 834 | 2 | US-08-677-734A-10 | Sequence 10, Appli |
| 34 | 108.5 | 3.8 | 1299 | 4 | US-08-460-900C-62 | Sequence 62, Appli |
| 35 | 107 | 3.8 | 1285 | 2 | US-08-540-406-6 | Sequence 6, Appli |
| 36 | 107 | 3.8 | 1285 | 3 | US-08-656-055-6 | Sequence 6, Appli |
| 37 | 107 | 3.8 | 1285 | 4 | US-08-954-668-6 | Sequence 6, Appli |
| 38 | 107 | 3.8 | 1285 | 5 | PCT-US95-13233-6 | Sequence 6, Appli |
| 39 | 103.5 | 3.6 | 1286 | 4 | US-09-268-140-3 | Sequence 3, Appli |
| 40 | 103 | 3.6 | 548 | 3 | US-08-903-139B-28 | Sequence 28, Appli |
| 41 | 101 | 3.6 | 470 | 2 | US-08-724-394A-10 | Sequence 10, Appli |
| 42 | 99 | 3.5 | 455 | 1 | US-08-035-928-2 | Sequence 2, Appli |
| 43 | 98 | 3.4 | 547 | 3 | US-08-903-139B-7 | Sequence 7, Appli |
| 44 | 98 | 3.4 | 550 | 4 | US-08-637-823B-30 | Sequence 30, Appli |
| 45 | 97 | 3.4 | 366 | 1 | US-08-554-612C-50 | Sequence 50, Appli |

ALIGNMENTS

RESULT 1
US-08-501-572-3
; Sequence 3, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport of protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohy, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-501-572-3

SEQUENCE UNPARISON 'A'

Query Match 24.7%; Score 702; DB 3; Length 555;
Best Local Similarity 33.7%; Pred. No. 4.9e-65;
Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;

QY 5 DEVIATLGEWGPQRLIFF---LLSASIIPNGFNGSVVFLAGTPEHRCRVPDANLS--59



✓

```

RESULT      2
US-09-040-444-3
: Sequence 3, Application US/09040444
: Patent No. 6063786
: GENERAL INFORMATION:
: APPLICANT: Koepsell, Hermann
: APPLICANT: Grudeman, Dirk
: APPLICANT: Gorboulev, Valentin
: TITLE OF INVENTION: Transport Of Cationic Xenobiotics And
: TITLE OF INVENTION: protein Which Effects The
: TITLE OF INVENTION: Transport Of Cationic Xenobiotics And
: TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
: TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.

```

COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,444
FILING DATE: March 19, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Steven P
REGISTRATION NUMBER: 41,225
REFERENCE/DOCKET NUMBER: 2481.1453-01
COMMUNICATION INFORMATION:

Query Match 24.7%; Score 702; DB 3; Length 555;
Best Local Similarity 33.7%; Pred. No. 4.9e-65;
Matches 187; Conservative 93; Mismatches 229; Indels 4

RESULT 3
US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:

APPLICANT: Koepsell, Heimgartner
APPLICANT: Grundeman, Dirk
APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport protein which effects the
TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSED: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington

1000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: * August 16, 2001, 13:55:18 ; Search time 20.69 Seconds
(without alignments)
548.346 Million cell updates/sec

Title: US-09-521-195-1
Perfect score: 2845
Sequence: 1 MRDYDEVIAFLGEGWPFQRL.....KKTRDSMETEENPKVLITAF 551

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap: *
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap: *
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap: *
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap: *
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------------|-------------------|
| 1 | 702 | 24.7 | 555 | 3 US-08-501-572-3 | Sequence 3, Appli |
| 2 | 702 | 24.7 | 555 | 3 US-09-040-444-3 | Sequence 3, Appli |
| 3 | 687.5 | 24.2 | 556 | 3 US-08-501-572-1 | Sequence 1, Appli |
| 4 | 687.5 | 24.2 | 556 | 3 US-09-040-444-1 | Sequence 1, Appli |
| 5 | 680 | 23.9 | 553 | 3 US-08-501-572-2 | Sequence 2, Appli |
| 6 | 680 | 23.9 | 553 | 3 US-09-040-444-2 | Sequence 2, Appli |
| 7 | 630 | 22.1 | 537 | 2 US-08-647-397-2 | Sequence 5, Appli |
| 8 | 234.5 | 8.2 | 494 | 2 US-09-031-392-5 | Sequence 5, Appli |
| 9 | 234.5 | 8.2 | 494 | 2 US-09-299-549-5 | Sequence 5, Appli |
| 10 | 222 | 7.8 | 492 | 2 US-08-355-844-3 | Sequence 3, Appli |
| 11 | 222 | 7.8 | 492 | 5 PCT-US95-16126-3 | Sequence 3, Appli |
| 12 | 216 | 7.6 | 493 | 4 US-09-031-392-10 | Sequence 10, Appl |
| 13 | 216 | 7.6 | 493 | 4 US-09-299-549-10 | Sequence 10, Appl |
| 14 | 213.5 | 7.5 | 524 | 2 US-08-928-692-12 | Sequence 12, Appl |
| 15 | 196 | 6.9 | 500 | 2 US-09-031-392-7 | Sequence 7, Appli |
| 16 | 196 | 6.9 | 500 | 4 US-09-299-549-7 | Sequence 7, Appli |
| 17 | 181 | 6.4 | 509 | 2 US-09-031-392-6 | Sequence 6, Appli |
| 18 | 181 | 6.4 | 509 | 4 US-09-299-549-6 | Sequence 6, Appli |
| 19 | 174.5 | 6.1 | 584 | 2 US-08-928-692-13 | Sequence 13, Appl |
| 20 | 172 | 6.0 | 383 | 2 US-09-031-392-3 | Sequence 3, Appli |
| 21 | 172 | 6.0 | 383 | 4 US-09-299-549-3 | Sequence 3, Appli |
| 22 | 169.5 | 6.0 | 488 | 2 US-08-928-692-11 | Sequence 11, Appl |
| 23 | 168.5 | 5.9 | 488 | 2 US-08-928-692-10 | Sequence 10, Appl |
| 24 | 162 | 5.7 | 563 | 4 US-09-031-392-2 | Sequence 2, Appli |
| 25 | 162 | 5.7 | 563 | 4 US-09-299-549-2 | Sequence 2, Appli |
| 26 | 158.5 | 5.6 | 109 | 2 US-08-647-397-4 | Sequence 4, Appli |
| 27 | 156.5 | 5.5 | 534 | 2 US-09-031-392-4 | Sequence 4, Appli |

ALIGNMENTS

RESULT 1
US-08-501-572-3
; Sequence 3, Application US/08501572
; Patent No. 6063623

; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann

; APPLICANT: Grundeman, Dirk

; APPLICANT: Gorboulev, Valentin

; TITLE OF INVENTION: Transport protein Which Effects The

; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Carrett & Dunner

; STREET: 1300 I Street, N.W., Suite 700

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572

; FILING DATE:
; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M

; REGISTRATION NUMBER: 35,391

; REFERENCE/DOCKET NUMBER: 02481.1453-00000

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000

; TELEFAX: (202)408-4400

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 555 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-501-572-3

Query Match 24.7%; Score 702; DB 3; Length 555;

Best Local Similarity 33.7%; Pred. No. 4.9e-65;

Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;

QY 5 DEVIAFLGEGWPFQRLFF---LLSASLIPIGFMGMSVFLAGTPEHRCRVPDANLUS-- 59

Db 6 DVLEHGGEFFQKQMFLLALLSATAP---IYGVIFLGFTHDRCRSPGVAELSLR 62
QY 60 SAWR-----NNSVPLRLRDGVPVHSCRYRLA-TIANFSALGLEPGDVLGQLEQESC 113
Db 63 CGWSAEELNYTVPGPGAGEASPROCRRYEVDWNOSTFDCVDPDLASLDTNRSRLPLGPC 122
QY 114 LDGWEFSQDYLSTVVTENLVNWCENWVPLTTSFFVGVGLIGSVQSLSDRFGRKNVL 173
Db 123 RDGWY--ETPGSSIVTEFNLCANSMWLDLFQSSVNVGVFFIGSMISGYIADRFGRKLCL 180
QY 174 FATMAVQTGFSFIQFISISWEMFTVLFVIVGMQOISNVVAFILGTEILGKSVRIIFSTL 233
Db 181 LTTVLINAAAGVLMASPTYTWMILFRILQGLVSKAGWLGILITFEVGGRYR---RTV 237
QY 234 GVC--TFPFAVGVMLLPLFAFYFIRDRMMLLALTVPGLVCLVPLWFWFIPESPRWLISQRFR 291
Db 238 GIFYQVAYTVGLVLVAGVAYALPHWRWLQFTVALPNFFLLYYWCIPESPRWLISQNKNA 297
QY 292 EAEDIIQKAAMNNTAVPAVI-----FDSVEELNPLKQKAFILDLFTRNIAIMTMS 345
Db 298 EAMRIKIHAKNKGLSLPASLQRLLEBETGKKNP-----SFLDLVTPQIRKHTMIL 351
QY 346 LLLWMLTSVGYFALSADPNLHGD--AYLNCFLSALIEIPAYITAWLLTLPRYIIAAV 404
Db 352 MYNFTSSVLYQGLIMHM--GLAGDNIYLDFFYSALVEFPAFMIILIIDRIGRRYPWAAS 410
QY 405 LFWGGGVLLFIQLVDPVDFYFISGLVMLGKGITSAFMSLVYFAELVPTLVNMAVGV 464
Db 411 NMVAGAACLASVFIIPGDLQWLKIIISCLGRMGITMAYEIVCLVNAELVPTFIRNLGVHIC 470
QY 465 STASRVGSIIPFYVYLGAYNRM-----LPYIVMGSLTVLIGITLFPFESLGMTLPET 518
Db 471 SSMCDIGGIITPELVY-----RLTNWLELPLMVFGVLVAGGLALLLPETKGALEPET 525
QY 519 LEQMOKVKWFRSGKK 533
Db 526 IEAEENMQRPRKNKE 540

RESULT 2

US-09-040-444-3
Sequence 3, Application US/09040444
Patent No. 6063766
GENERAL INFORMATION:
APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport protein which effects the
TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,444
FILING DATE: March 18, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Steven P
REGISTRATION NUMBER: 41,225
REFERENCE/DOCKET NUMBER: 2481.1453-01
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-040-444-3

Query Match 24.7%; Score 702; DB 3; Length 555;
Best Local Similarity 33.7%; Pred. No. 4.9e-65;
Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;

QY 5 DEVIATFGEWGPQRLIFF---LLSASIPNGFNGSVVFLAGTPEHRCRVPDAANLS-- 59
Db 6 DVLEHGGEFFQKQMFLLALLSATAP---IYGVIFLGFTHDRCRSPGVAELSLR 62
QY 60 SAWR-----NNSVPLRLRDGVPVHSCRYRLA-TIANFSALGLEPGDVLGQLEQESC 113
Db 63 CGWSAEELNYTVPGPGAGEASPROCRRYEVDWNOSTFDCVDPDLASLDTNRSRLPLGPC 122
QY 114 LDGWEFSQDYLSTVVTENLVNWCENWVPLTTSFFVGVGLIGSVQSLSDRFGRKNVL 173
Db 123 RDGWY--ETPGSSIVTEFNLCANSMWLDLFQSSVNVGVFFIGSMISGYIADRFGRKLCL 180
QY 174 FATMAVQTGFSFIQFISISWEMFTVLFVIVGMQOISNVVAFILGTEILGKSVRIIFSTL 233
Db 181 LTTVLINAAAGVLMASPTYTWMILFRILQGLVSKAGWLGILITFEVGGRYR---RTV 237
QY 234 GVC--TFPFAVGVMLLPLFAFYFIRDRMMLLALTVPGLVCLVPLWFWFIPESPRWLISQRFR 291
Db 238 GIFYQVAYTVGLVLVAGVAYALPHWRWLQFTVALPNFFLLYYWCIPESPRWLISQNKNA 297
QY 292 EAEDIIQKAAMNNTAVPAVI-----FDSVEELNPLKQKAFILDLFTRNIAIMTMS 345
Db 298 EAMRIKIHAKNKGLSLPASLQRLLEBETGKKNP-----SFLDLVTPQIRKHTMIL 351
QY 346 LLLWMLTSVGYFALSADPNLHGD--AYLNCFLSALIEIPAYITAWLLTLPRYIIAAV 404
Db 352 MYNFTSSVLYQGLIMHM--GLAGDNIYLDFFYSALVEFPAFMIILIIDRIGRRYPWAAS 410
QY 405 LFWGGGVLLFIQLVDPVDFYFISGLVMLGKGITSAFMSLVYFAELVPTLVNMAVGV 464
Db 411 NMVAGAACLASVFIIPGDLQWLKIIISCLGRMGITMAYEIVCLVNAELVPTFIRNLGVHIC 470
QY 465 STASRVGSIIPFYVYLGAYNRM-----LPYIVMGSLTVLIGITLFPFESLGMTLPET 518
Db 471 SSMCDIGGIITPELVY-----RLTNWLELPLMVFGVLVAGGLALLLPETKGALEPET 525
QY 519 LEQMOKVKWFRSGKK 533
Db 526 IEAEENMQRPRKNKE 540

RESULT 3

US-08-501-572-1
Sequence 1, Application US/08501572
Patent No. 6063623
GENERAL INFORMATION:
APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport protein which effects the
TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington

STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: US/08/501,572
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Toohy, Kimberlin M
REGISTRATION NUMBER: 35,391
REFERENCE/DOCKET NUMBER: 02481.1453-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-501-572-1

Query Match 24.28; Score 687.5; DB 3; Length 556;
Best Local Similarity 33.28; Pred. No. 1.6e-63;
Matches 184; Conservative 88; Mismatches 230; Indels 53; Gaps 11;
QY 1 MRDYDEVIAFLGEMGFQRLIFLL---SASIIPIGFMGMSVFLAGTPHRCRVPDAA 57
Db 1 MPTVDVLEQGEFGWFQKQAFLLCLLSASLAP---IYGVIFLGTGPGHYCQNPQVAE 57
QY 58 LSS--AWR-----NNSVP-LRLRGREVPHSCSYRL-----ATIANFSA 95
Db 58 LSQRGWSQAEBLNTVPGLGPSDEASFLSQCMRYEVDWNGSTLDCVDP LSSLVANRSQL 117
QY 96 GLEPGRDVLDGLESCLDGWEFSODVYLSVVTENWLVCEDNWKP LTTSLFFVGVLL 155
Db 118 PLGP-----CEHGMY--DTPGSSIVTEFNLCVGDANKVDLQSCVNLGFFL 162
QY 156 GSFVSQSLSDRGKRVNLFATMAVQTGFSFLOIFSISWEMFTVLFIIVGMQISNYVAF 215
Db 163 GSLVVGVIADRGKRLCLLVTLVTSVSGVLTAVAPDYSMLLFRLLQGMVSKGWSYSGY 222
QY 216 ILGTEILGKSVR---IIFSTLGVCTFFAVGYMLLPLFAFIRDRWMLLLALTPVGLCV 271
Db 223 TLITEFVSGYRRTTALLYQMA-----FTVGLVGLAGVAYAIPOWRMLQLAVSLTFL 277
QY 272 PLWFTIPESPRWLISORRPREAEDIIOKAAKNMNTAVPAVIFDSVEELNPLKQOKAFILD 331
Db 278 LYWFVPEPRWLLSQKRTTRAVRIMEQIAQKNGVPPADLKMLCLEDEASEKSPSFAD 337
QY 332 LFRTRNIAINTMSLLMMLTSGVGFALSODAPNLHGDAYNCLFSLALIEPAITAWLL 391
Db 338 LFRTPNLRKHTVIMLYMFSCAVLYQGLIMHVGTAGANLYLDFFYSLSVPEPAFIILVT 397
QY 392 LRTLPRYIIAALFWGGGVLFIQLVPDYYFLSIGLVMLGKFGITSASFMSLVYETAEL 451
Db 398 IDRIGRIYPIAASNLVTGAACLLMIFIPHELHMLNVTIACLRGMGATVLEMCVLVNAEL 457
QY 452 YPTLVNMAVGTSTASRVGSIAPVYV-LGAYNRMPLPYVMGSLTILGIFTIFFPES 510
Db 458 YPTFIRNLGMVCSALCDLGGITFPNVRIMEVWQALPLILFGVLGITAGAMILLIPET 517
QY 511 LGMTLPETLEQMQV 525
Db 518 KGVALPETIEAENL 532

RESULT 4
US-09-040-444-1
Sequence 1, Application US/09040444
Patent No. 6063766
GENERAL INFORMATION:
APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals,
TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals,
TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,444
FILING DATE: March 18, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Steven P
REGISTRATION NUMBER: 41,225
REFERENCE/DOCKET NUMBER: 2481.1453-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-040-444-1

Query Match 24.28; Score 687.5; DB 3; Length 556;
Best Local Similarity 33.28; Pred. No. 1.6e-63;
Matches 184; Conservative 88; Mismatches 230; Indels 53; Gaps 11;
QY 1 MRDYDEVIAFLGEMGFQRLIFLL---SASIIPIGFMGMSVFLAGTPHRCRVPDAA 57
Db 1 MPTVDVLEQGEFGWFQKQAFLLCLLSASLAP---IYGVIFLGTGPGHYCQNPQVAE 57
QY 58 LSS--AWR-----NNSVP-LRLRGREVPHSCSYRL-----ATIANFSA 95
Db 58 LSQRGWSQAEBLNTVPGLGPSDEASFLSQCMRYEVDWNGSTLDCVDP LSSLVANRSQL 117
QY 96 GLEPGRDVLDGLESCLDGWEFSODVYLSVVTENWLVCEDNWKP LTTSLFFVGVLL 155
Db 118 PLGP-----CEHGMY--DTPGSSIVTEFNLCVGDANKVDLQSCVNLGFFL 162
QY 156 GSFVSQSLSDRGKRVNLFATMAVQTGFSFLOIFSISWEMFTVLFIIVGMQISNYVAF 215
Db 163 GSLVVGVIADRGKRLCLLVTLVTSVSGVLTAVAPDYSMLLFRLLQGMVSKGWSYSGY 222
QY 216 ILGTEILGKSVR---IIFSTLGVCTFFAVGYMLLPLFAFIRDRWMLLLALTPVGLCV 271
Db 223 TLITEFVSGYRRTTALLYQMA-----FTVGLVGLAGVAYAIPOWRMLQLAVSLTFL 277
QY 272 PLWFTIPESPRWLISORRPREAEDIIOKAAKNMNTAVPAVIFDSVEELNPLKQOKAFILD 331
Db 278 LYWFVPEPRWLLSQKRTTRAVRIMEQIAQKNGVPPADLKMLCLEDEASEKSPSFAD 337

QY 332 LFRTRNIAIMTMSLLMLTSSVGFALSADPNLHGDVYLNCFLSALIEIPAYITAWLL 391
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 338 LFRTPNKRHTVILMLWFSCAVLYQGLIMHVGATGANLYLDFYSSLYVEEPAFIILVT 397
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 392 LRTLPRIIAAVLFWGGVLLFTQLVPDYVYFSLISGLVLMGKFGITSAFMSLYVTAEL 451
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 398 IDRIIGRIPIAASNLVTGAACLLMIFIPHELHMLNVTLACLRMGATVLEWCLVNAEL 457
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 452 YPTLRNMAVGTSTASRVSIAPYEVY-LGAYNRMLPYIVMGSITVLIGITFTFPFES 510
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 458 YPTLRNMGVMVSCALCDLGIFTPFVFRMLMEVWQALPILFSLGLTAGAMTLLPET 517
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 511 LGMTLPETLEOMQKV 525
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 518 KGVALPETIEEAHL 532
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

RESULT 5
US-08-501-572-2
; Sequence 2, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohy, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-501-572-2

Query Match 23.9%; Score 680; DB 3; Length 553;
Best Local Similarity 33.4%; Pred. No. 9.9e-63;
Matches 192; Conservative 85; Mismatches 228; Indels 70; Gaps 15;
QY 1 MRDYDEVIATLFGKQFPQR---LIFFLSASIIIPNGFNGSVVFLAGTPEHRCRVPDAAN 57
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 1 MPTVDDILEQVGSQWQKQAFLLCLLSAFAF---ICVGIYVFLGTPDHCQSPGVAE 57
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 58 LSS---AWR-----NNSVPLRLRQREVPHSCSYR-----LATTA-NFSALG 96
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

Db 58 LSRCGWSPAEELNYTPGLGPAGEAFLGQCRRYEVDWNOQSALSQVDPLASLATNRSHLP 117
QY 97 LEPGRVDLGLQEGESCDQWESQDQVYLSVTWNLVCEDNWKVPLTTSLSFFVGVLLG 156
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 118 LGP-----CODGWY--DTPGSSIVTEENLVACADSKLDDLQSCNLNAGFFFG 162
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 157 SFVSGQLSDRGRKNVLFATMAVQTGFSFIQFISINWEMFTVLFVIVMGQISNYVVAFI 216
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 163 SLGVGYFADREFGRKLCLLGLTVLNAVSGVLMFSPNYMSMLLFRLLQGLSVKGNMAGYT 222
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 217 LGTELG----KSVRIIFSTLGVCTRFVAVGYMLLFAFYFIRDRMLLLALTVPGLVCP 272
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 223 LITEFVSGSRRTVAIMQMA----FTVGLVALTGLAYALPHWRMLQLAVSLPFLFL 277
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 273 LWFIPESRWLISQRRFREAEIDIIKAAKMNNTAVPAVI-----FDSVEELNPLKQOK 326
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 278 YYWCVPESPRWLLSKRNTAEIKIMDHIAOKNGKLPADLKMLSLEEDVTEKLS- 332
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 327 AFILDLEFRNIAIMTMSLLMLTSSVGFALSADPNLHGDVYLNCFLSALIEIPAYI 386
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 333 -SFADLFRTPRLKRTFILMFTDTSVLQGLIHHMGATSGNLYLDFLYSALVEIPGAF 391
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 387 TAWLLRLTPRRYIIAAVLFWGGVLLFIQLVPDYVYFSLISGLVLMGKFGITSAFMSLYV 446
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 392 IALITIDRVGRIPYMAVSNLLAGAACLVIFISP-DLHNLNIIIMCVGRGITIAIQMICL 450
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 447 FTABLYPTLRNMAVGTSTASRVSIAPYEVY-LGAYNRMLPYIVMGSITVLIGITFTL 505
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 451 VNAELYPTFVRNLRVMVSCSLCDIGGIITPFIIVFRLEWQALPLILFAVLGLAAGVTL 510
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 506 FPFESLGMTLPETLEOMQKVWFRSKKTRDSMET 540
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 511 LLPETKDALPETMKDAENL----GRKAKPKENT 540
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

RESULT 6
US-09-040-444-2
; Sequence 2, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids

```

;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-040-444-2

Query Match      23.9%; Score 680; DB 3; Length 553;
Best Local Similarity 33.4%; Pred. No. 9.9e-63;
Matches 192; Conservative 85; Mismatches 228; Indels 70; Gaps 15;

QY 1 MRDVEIAFLGEGWPFQRLIFLLSASIIIPNGFMGMSVFLAGTPEHRCRVDPDAAN 57
Db 1 MPTVDDILLEGWESGWFKQAFLILLLSAFAP--ICVGIVFLGFTPDHHCQSGVAE 57
QY 58 LSS--AMR-----NNSVPLRLRDGREVPHSCSRY-----LATIA-NFSALG 96
Db 58 LSQRCGSPAEELNNTVPGGLGAGEAFUGQRCRRYEDVWQNSALSVDPLASLATNRSLP 117
QY 97 LEPRDVLGQLESCDGFESQDYLSTVVTENNLYCEDNWKVPLTTSLEFFVGVLG 156
Db 118 LGP-----CQDGVVY--DTPGSSIVTEFNLCVADSKKLDLFOSCLNAGFFG 162
QY 157 SFVSGQLSDRFGKRNLFATMAVGTGFSFLQIFSIEMETVLFVIVGMGOISNVVAFI 216
Db 163 SLGVGFADREGRKLCLLGTIVNVAHSVLMASFNTYMSMLLFRLLQGLVSKGNWAGYT 222
QY 217 LGTEILG----KSVRIIFSTLGVCTFFAVGMYMLLPLFAYFIRDRMRLLLALTVPGLVCP 272
Db 223 LITEVSGSRRTVAIMYQMA----FTVGLVALTGLAYALPHWRWLQAVSLPTFLFL 277
QY 273 LWWFIPESPRMLISQRRPREAEDIIQRAKNNTAVPAVI-----FDSVEELNPLKQOK 326
Db 278 YYWCVPESPRMLLSQKRNTAIIKMDHTAQNKKLPPADLKMLSLEEDVTEKLSF 332
QY 327 AFILDLFRTRNIAITIMSLLLMWTSGVGFALSIDAPNLHGDAYLNCFLSALTEIPAYI 386
Db 333 -SFADLFTPLRKRKFTILMLTWDTSVLYOGLIHWGATSGNLYLDFLYSALVEIPAF 391
QY 387 TAWLLRLTPRYIIAIVLFWGGVLLFIQLVPVDYIYFLSGLVLMKFGTTSAFSMLYV 446
Db 392 IALITIDRVGRIPMAVSNLLAGAACLVIFISP-DLHWLNIIMCVGRMGITIAIQICL 450
QY 447 FTAELYPILVRMAGVVTASRVSIIATPFVY-LGAYNRMLPIYVMSLTVLIGITL 505
Db 451 VNAELYPFVRNLRVMVSSICLDIGGIITPFVFLREVWQALPLILFAVLGLAAGVTL 510
QY 506 FPPEISGLMTLPETLEQMKVKWFRSGKTRDSMET 540
Db 511 LLPETKGDALPETMKDAENL-----GRKAKPKENT 540

RESULT 7
US-08-647-397-2
; Sequence 2, Application US/08647397
; Patent No. 5972702
; GENERAL INFORMATION:
; APPLICANT: Beier, David R.
; APPLICANT: Brady, Kevin P.
; TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,397
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7048
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-647-397-2

Query Match      22.1%; Score 630; DB 2; Length 537;
Best Local Similarity 30.0%; Pred. No. 1.7e-57;
Matches 165; Conservative 102; Mismatches 245; Indels 38; Gaps 12;

QY 4 YDEVIAFLGEGWPFQRLIFLLSASIIIPNGFMGMSVFLAGTPEHRCRVDPDAANSSAWR 63
Db 3 FSEILDRVSGMGPFQYLHVTLALPILGIANHNLLQIFATTTPDHCRPPPNASL-EPW- 60
QY 64 NNSVPLRLRDGREVPHSCSRY---RLATIANFSALEGRDVLGQLESCDGFES 120
Db 61 --VLPLG-PNGK--PEKCLRFVHLNPNASLNDTQGATEP-----CLDGWIYN 102
QY 121 QDYVLSVTVTENNLYCEDNWKVPLTTSLEFFVGVLGFSVSGQLSDRFGKRNLFATMAVQ 180
Db 103 ST--RTVITENDVLGCGSNKLEMAQSVFMAGILVGGPVFSGELSDRFGKRPILTWSYLLL 160
QY 181 TGFSEFLQIFSIEMETVLFVIVGMGOISNVVAFILGTEILGKSVRIIFS-TLGVCTFF 239
Db 161 AASGSSNAFSPSLTVYMIFFLGCSCISGISLSTIILNVWVPTSTRAISSTTIGVC--Y 218
QY 240 AVGYMLLPLFAYFIRDRMRLLLALTVPGLVCPDLMWFIPESPRMLISQRRPREAEDIIQK 299
Db 219 TIGQFILPLGLAYAVPQWRWLQLSVSAAFFIFSLLSWVPESIRWLVSGLKFSKALKTLQR 278
QY 300 AAKMN-----NTAPAVIFDSVEELNPLKQOKAFITLDLFRTRNIAITIMSLLLMWT 352
Db 279 VATFNGKKEGKELTVEELAFNLOKDTISAK-VKYGSLDLFRVSIILRRVTFCLSLAWFAT 337
QY 353 SVGYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAIVLFWGGVLL 412
Db 338 GFAYISLAMGVVEFGVNIYILQIIFGGVDIPAKFITILSISYLGRRITQCFLLILAGVAI 397
QY 413 LFIQLVPVDYIYFLSGLVLMKFGTTSAFSMLYVFAELYPILVRMAGVVTASRVS 472
Db 398 LALIFVSSEMQLLRTALAVFGKCLSGFSCLFLYSELPTVLRGTGMSISINWARGVS 457
QY 473 IAPYFVYLGAYNRMLPIYVMSLTVLIGITLFFFPFSLGMLTLPETLEQMKVKWFRSGK 532
Db 458 MIAPLVKITGELQPFIPNVIFWTMLLGGSAAPFLETLNRPLEPIEDIQ--DWQQTQK 515
QY 533 KTRDSMETEE 542
Db 516 KTKQEPKAEK 525

RESULT 8
US-09-031-392-5
; Sequence 5, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
```


QY 530 SGKTRDSMETE 541
|||
Db 503 SGSARPKAAVE 514

RESULT 15

US-09-031-392-7
; Sequence 7, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-031-392-7

Query Match 6.9%; Score 196; DB 2; Length 500;
Best Local Similarity 24.8%; Pred. No. 4.9e-12;
Matches 122; Conservative 78; Mismatches 175; Indels 116; Gaps 28;
QY 118 EFSQDYLSTVVTENLVCEDNWKVPLTSLFFVGLGSEFVSGQLSDRFGKKNVLFATM 177
|||
Db 60 EPMEDFPLTL-----WSV--TVSMFPEGFGISLLVGLVPLVYKFKRGKALL--- 103
QY 178 AVQTGFSFIQIFSI-----SWEMFTVLVIVGM-GQISNVVVAIFILGTEILG 223
|||
Db 104 -----FNNIFSVIPALMGCSRVSFELIISRLVGLGICAGVSSNVVPMYLG-ELAP 155
QY 224 KSVRIIFSTLGYC--TFPAVGVMILPLFAY-----FIRDWRMLLALTVPGLVCLVPLWVF 276
|||
Db 156 KNLR--GALGVVPQLFTVIGLVAQIFGLRNLNLANVDGWFILLGLTGVPAALQLLLLPF 212
QY 277 IPESRWLISORRPFHEADITQKAKMNTAVPAVIFDSVE-ELNPLKQ-----OKA---- 327
|||||
Db 213 FPESRYLLIQKK-----DEAAKRALQTLRG--WDSVDREVAEIRQDEAEKAAAGFI 263
QY 328 FTLDLFRTRNTAINTMSLLIWM-----LTSVGYFA-----LSLDAPNLHGDAYLNCFL 376
:|
Db 264 SVLKLFRMRSLR-WQLLSIIVMGQQLSGVNAIYYADQIYLSAGVPEEHVQ-YVTAGT 321
QY 377 SALIEIPAYITAWLRLTPRRYIAAFLFWGGGVLLFIQLVPVDYVYFLSLGLVMLGKFG 436

Db 322 GA-VNVVMTFCAVFVVELLGRLL-----LLLGFSICLIACCCVLTAALALQDTVSWMPYIS 376
QY 437 ITSAPSMLYVFTAEELYPTLVNRNAVGVTTSTASR-----VG-----SIIAPYFVY 480
|||
Db 377 IVCVIS--YVIGHALGSPSPALLITIFLQSSRPSAFVWGVGVHWSNFTVGLIFPPTQE 434
QY 481 -LGAYNRMLPYIVMGSLVNLIGIFT-LFFPESLGMT---LPETLEQMKYKWFPSGKTR 535
|||
Db 435 GLGPYS----FIVFAVICLITTIYIFLIVPETKAKTFIEINQIFTKMKVSEVYPEKE-- 488
QY 536 DSMETEENPKV 546
|||
Db 489 ---ELKELPPV 496

Search completed: August 16, 2001, 13:58:52
Job time: 214 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:54:58 ; Search time 36.39 Seconds

(without alignments)

917.939 Million cell updates/sec

Title: US-09-521-195-1

Perfect score: 2845

Sequence: 1 MRDYDEVIAFLGEGWGFQRL.....KKTRDSMETENPKVLITAF 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-------------|--------------------|
| 1 | 2845 | 100.0 | 551 | 20 AAY01649 | A protein with cat |
| 2 | 2470 | 86.8 | 553 | 20 AAY01851 | A protein with cat |
| 3 | 2470 | 86.8 | 553 | 21 AAB20579 | Mouse OCTN1 amino |
| 4 | 2218 | 78.0 | 557 | 20 AAY01650 | A protein with cat |
| 5 | 2218 | 78.0 | 557 | 21 AAY83929 | Human carnitine tr |
| 6 | 2164 | 76.1 | 557 | 20 AAY01652 | A protein with cat |
| 7 | 2164 | 76.1 | 557 | 21 AAB20580 | Mouse OCTN2 amino |
| 8 | 2164 | 76.1 | 557 | 21 AAY83930 | Mouse carnitine tr |
| 9 | 2034.5 | 71.5 | 564 | 21 AAB20578 | Mouse OCTN3 protei |
| 10 | 779.5 | 27.4 | 560 | 21 AAB43038 | Human ORFX ORF2802 |
| 11 | 698 | 24.5 | 540 | 22 AAB49401 | Murine organic ani |

| | | | | | |
|----|-------|------|-----|-------------|--------------------|
| 12 | 685.5 | 24.1 | 554 | 19 AAW64538 | Human liver cell c |
| 13 | 682 | 24.0 | 535 | 21 AAY51249 | Rat liver anion tr |
| 14 | 681.5 | 24.0 | 556 | 17 AAR77676 | Rat OCT-1 protein. |
| 15 | 653.5 | 23.0 | 548 | 21 AAB08823 | A human organic an |
| 16 | 651 | 22.9 | 542 | 21 AAY92902 | Human cerebral org |
| 17 | 650 | 22.8 | 551 | 21 AAB08824 | A human organic an |
| 18 | 637 | 22.4 | 536 | 21 AAY92903 | Rat cerebral organ |
| 19 | 636 | 22.4 | 561 | 18 AAW44196 | Human osteoclast t |
| 20 | 630 | 22.1 | 537 | 18 AAW44195 | Mouse osteoclast t |
| 21 | 624 | 21.9 | 607 | 21 AAB12131 | Hydrophobic domain |
| 22 | 606.5 | 21.3 | 545 | 22 AAB36553 | Mouse organic anio |
| 23 | 597.5 | 21.0 | 551 | 20 AAW84888 | Rat organic anion |
| 24 | 586 | 20.6 | 550 | 21 AAY44278 | Human organic anio |
| 25 | 583.5 | 20.5 | 563 | 20 AAW84889 | Human organic anio |
| 26 | 564 | 19.8 | 550 | 22 AAB69091 | Human organic anio |
| 27 | 417 | 14.7 | 480 | 21 AAG28044 | Arabidopsis thalia |
| 28 | 417 | 14.7 | 483 | 21 AAG28043 | Arabidopsis thalia |
| 29 | 412.5 | 14.5 | 521 | 21 AAG30885 | Arabidopsis thalia |
| 30 | 387.5 | 13.6 | 397 | 21 AAG30886 | Arabidopsis thalia |
| 31 | 382.5 | 13.4 | 515 | 21 AAG42522 | Arabidopsis thalia |
| 32 | 341 | 12.0 | 360 | 21 AAG30887 | Arabidopsis thalia |
| 33 | 325.5 | 11.4 | 378 | 21 AAG42523 | Arabidopsis thalia |
| 34 | 312.5 | 11.0 | 359 | 21 AAG42524 | Arabidopsis thalia |
| 35 | 305.5 | 10.7 | 439 | 22 AAB76766 | Corynebacterium gl |
| 36 | 303 | 10.7 | 297 | 21 AAG28045 | Arabidopsis thalia |
| 37 | 302.5 | 10.6 | 548 | 21 AAY44633 | Human organic cati |
| 38 | 277 | 9.7 | 339 | 21 AAB19406 | Amino acid sequenc |
| 39 | 272.5 | 9.6 | 231 | 17 AAR77677 | Partial human OCT- |
| 40 | 267 | 9.4 | 520 | 20 AAY06115 | Human organic cati |
| 41 | 267 | 9.4 | 520 | 21 AAY56881 | Human saccharide-t |
| 42 | 267 | 9.4 | 520 | 21 AAY53009 | Human secreted pro |
| 43 | 247 | 8.7 | 530 | 21 AAB29626 | Cat flea HMT synap |
| 44 | 243 | 8.5 | 399 | 22 AAB76717 | Corynebacterium gl |
| 45 | 243 | 8.5 | 492 | 12 AAR11360 | Glucose Transporte |

ALIGNMENTS

RESULT 1

AA1980

ID AAY01649 standard; Protein; 551 AA.

AC AAY01649;

XX AAY01649;

XX 23-JUN-1999 (first entry)

DT A protein with cation transporting activity.

DE Organic cation transporter; OCT1; OCT2; drug development; fatty liver;

KW heart disease; cancer; anti-tumour drug; anticancer drug.

XX Homo sapiens.

OS WO9913072-A1.

PN 18-MAR-1999.

PD 07-SEP-1998; 98WO-JP04009.

PF 20-MAY-1998; 98JP-0156660.

PR 08-SEP-1997; 97JP-0260972.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Nezu J, Oku A;

XX WPI: 1999-215062/18.

XX N-PSDB; AAX26879.

XX Genes homologous with organic cation transporters OCT1 and OCT2,

PT useful in design of new drugs for treatment of diseases due to

PT abnormality of the transporter functions

```

XX PS Claim 1; Page 41-45; 97pp; Japanese.
XX CC The present sequence represents a protein with cation transporting
XX CC activity. The genes are significantly homologous with organic cation
XX CC transporters OCT1 and OCT2. The genes may be used in drug development,
XX CC particularly in the treatment of diseases due to abnormality of the
XX CC organic cation transporter functions e.g. fatty liver, heart diseases
XX CC and cancers, by controlling such as by inhibition or activation.
XX CC Administration of anti-tumour and anticancer drugs in combination with
XX CC a transporter protein inhibiting agent allows the agents to penetrate
XX CC into the diseased cells to enhance the drug action.
XX SQ Sequence 551 AA;

Query Match 100.0%; Score 2845; DB 20; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.2e-281;
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDYDEVIAFLGEMGPFQRLIFFLLSASIIPIGNGSMVVFVLAGTPEHRCRVPDAANLSS 60
DB 1 mrdydeviaflgwgpfqrlifllsasiipngfngmsvfvlagtpehrcrvpdaanlss 60

QY 61 AWRNNSVPLRLRDGREVPHSCRYRLATIANFSAIGLEPGRDVDLQLEQESCLDGWEFS 120
DB 61 awrnnsvplrlrdgrevphscryrlatiansaiglepgdrvdldgleqescldgwefs 120

QY 121 ODVYLSVVTENLVCEDNWKPVTLSFFVGLGVSFVSGQLSDRGRKNVLFATMAVQ 180
DB 121 odvylstvtewnlvcednwkvptlsffvglgsvsfvsgqlsdrgrknvlfatmaq 180

QY 181 TGFSLQIFSIWEMFTVLFVIVGMGOISNYVAVFVILGTEILGKSVRIIFSTLGVCTFFA 240
DB 181 tgfslqifsiwemftvlfvivgmgoisnyvavfvlgtelilgksvriifstlgvctffa 240

QY 241 VGYMLPLPFAFIRDRMILLALTVPGLVPLWVFPESPRWLISQRRPREAEDIQKA 300
DB 241 vgymlplpfafrdrmillaltvpglvpplwvfpesprwlisqrrpreaediqlka 300

QY 301 AKMNTAVPAVIFDSVE--ELNPLKQKAFILDIFRTRNIAIMTMSLLIWLMLTSVGYFA 358
DB 301 akmntavpavifdsve--elnplkqkafildiftrtniaimtmslllmltsvgyfa 358

QY 361 LDAPNLHGDAYLNCFLSALIEIPAVITAWLLRLPRYIIAANLVFWGGVLLFIQLVPV 420
DB 361 ldapnlhgdaylncflsalieipavitawllrlpryiaaavlfwggvllfiqlvpv 420

QY 421 DYFSLIGLVMLGKFGITSFASMLYVFTAEALYPTLVNMAVGVTSAGRVGSIITAPFY 480
DB 421 dyfsliglvmlgkfgitsafsmlyvftaelyptlvnrmavgvtsagrvgsiitapfy 480

QY 481 LGAYNRLPYIVNGSLTVLIGITLFFPESIGMTLPETLEQMKVQKWPFRSKKTRDSMET 540
DB 481 lgaynrlpyivngsltvligitlffpesigmtlpetleqmkvkwpfrskktrdsmet 540

QY 541 EENPKVLITAF 551
DB 541 eenpkvlitaf 551

```

```

RESULT 2
AAY01651
ID AAY01651 standard; Protein; 553 AA.
XX AC AAY01651;
XX DT 23-JUN-1999 (first entry)
XX DE A protein with cation transporting activity.
XX KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KW heart disease; cancer; anti-tumour drug; anticancer drug.

```

```

XX OS Mus musculus.
XX PN WO9913072-A1.
XX PD 18-MAR-1999.
XX PF 07-SEP-1998; 98WO-JP04009.
XX PR 20-MAY-1998; 98JP-0156660.
XX PR 08-SEP-1997; 97JP-0260972.
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX PI Nezu J, Oku A;
XX DR WPI; 1999-215062/18.
XX DR N-PSDB; AAX26898.
XX PT Genes homologous with organic cation transporters OCT1 and OCT2,
XX PT useful in design of new drugs for treatment of diseases due to
XX PT abnormality of the transporter functions
XX PS Claim 1; Page 63-67; 97pp; Japanese.
XX CC The present sequence represents a protein with cation transporting
XX CC activity. The genes are significantly homologous with organic cation
XX CC transporters OCT1 and OCT2. The genes may be used in drug development,
XX CC particularly in the treatment of diseases due to abnormality of the
XX CC organic cation transporter functions e.g. fatty liver, heart diseases
XX CC and cancers, by controlling such as by inhibition or activation.
XX CC Administration of anti-tumour and anticancer drugs in combination with
XX CC a transporter protein inhibiting agent allows the agents to penetrate
XX CC into the diseased cells to enhance the drug action.
XX SQ Sequence 553 AA;

```

```

Query Match 86.8%; Score 2470; DB 20; Length 553;
Best Local Similarity 84.6%; Pred. No. 2.3e-243;
Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

QY 1 MRDYDEVIAFLGEMGPFQRLIFFLLSASIIPIGNGSMVVFVLAGTPEHRCRVPDAANLSS 60
DB 1 mrdydeviaflgwgpfqrlifllsasiipngfngmsvfvlagtpehrcrvpdtwnlss 60

QY 61 AWRNNSVPLRLRDGREVPHSCRYRLATIANFSAIGLEPGRDVDLQLEQESCLDGWEFS 120
DB 61 awrnnsipltkdgrgvqpsrtrylatiansaiglepgdrvdldgleqescldgweyd 120

QY 121 ODVYLSVVTENLVCEDNWKPVTLSFFVGLGVSFVSGQLSDRGRKNVLFATMAVQ 180
DB 121 odvylstvtewnlvcednwkvptlsffvglgsvsfvsgqlsdrgrknvlfatmaq 180

QY 181 TGFSLQIFSIWEMFTVLFVIVGMGOISNYVAVFVILGTEILGKSVRIIFSTLGVCTFFA 240
DB 181 tgfslqifstnwemftvlfvivgmgoisnyvavfvlgtelilgksvriifstlgvctffa 240

QY 241 VGYMLPLPFAFIRDRMILLALTVPGLVPLWVFPESPRWLISQRRPREAEDIQKA 300
DB 241 vgymlplpfafrdrmillaltvpglvpplwvfpesprwlisqrrpreaediqlka 300

QY 301 AKMNTAVPAVIFDSVE--ELNPLKQKAFILDIFRTRNIAIMTMSLLIWLMLTSVGYFA 358
DB 301 akmntavpavifdsve--elnplkqkafildiftrtniaimtmslllmltsvgyfa 358

QY 359 LSDAPNLHGDAYLNCFLSALIEIPAVITAWLLRLPRYIIAANLVFWGGVLLFIQLV 418
DB 361 lslnvpnlhgdaylncflsalieipavitawllrlpryiaaavlfwggvllfiqlv 420

QY 419 PDYFSLIGLVMLGKFGITSFASMLYVFTAEALYPTLVNMAVGVTSAGRVGSIITAPFY 478
DB 421 pedyfsliglvmlgkfgitsafsmlyvftaelyptlvnrmavgvtsagrvgsiitapfy 480

```

Qy 479 VYLGAYNRMLPYIVMGSLTVLIGIFTLFFPESLGMTLPETLEOMQVKVWFRSCKKTRDSM 538
Db 481 vylgaynrllpyilmgsltlvligitliffesgvtlpenleqmkgvrgfcgkktstvsv 540

Qy 539 ETEENPKVLITAF 551
Db 541 dreespkvlitaf 553

RESULT 3
AAB20579 AAB20579 standard; Protein: 553 AA.

AC AAB20579;
DT 11-DEC-2000 (first entry)

DE Mouse OCTN1 amino acid sequence.
KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
KW identification; regulator; carnitine transport.

OS Mus musculus.
PN WO200046368-A1.
XX 10-AUG-2000.

PF 04-FEB-2000; 2000WO-JP00619.
XX 05-FEB-1999; 99JP-0028406.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

PA Nezu J, Ose A;

PI WPI; 2000-586982/55.

XX Organic cation transporter gene OCTN3 expressed in testis for
XX identification of regulators of carnitine transport for use as drugs
XX Example 2; Fig 2; 58pp; Japanese.

XX The present invention describes a mouse organic cation transporter
CC protein (OCTN3). Also described are: (1) a method for screening
CC compounds for their ability to regulate the transport of an organic
CC cation into the cell, by generating a cell expressing OCTN3 at the
CC cell membrane, contacting with the compound and organic cation, and
CC observing the degree of transport of the organic cation; and (2) a
CC method for screening compounds for their ability to be transported into
CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
CC membrane, contacting with the compound and observing the degree of
CC transport of the compound. OCTN3 can be used for the identification of
CC regulators of the transport of organic cations (especially carnitine)
CC into cells by OCTN3, for use as drugs. The present sequence represents
CC the mouse OCTN1 amino acid sequence, which is used in an example from
CC the present invention.

XX Sequence 553 AA;

Query Match 86.8%; Score 2470; DB 21; Length 553;
Best Local Similarity 84.6%; Pred. No. 2.3e-243;
Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

Qy 1 MRDYDEVIAFLGEMGFQRLIFLLSASITPNGFNGMSVVFVLAGTPEHRCRVPDAANLSS 60
Db 1 mrdydeviafigewgpfqrliflllsasitpningmsvfvflagtpehrcrlvpdvtvlss 60

Qy 61 AWNNNSVPLRLDRGVRPHVSCSRYLATIANFSALGLEPGRDVLQGLQESCLDGWEFS 120
Db 61 swrnhslpletkgqvpsccrrylatiansamglepgqvdleqleqescldgweydw 120

Qy 121 QDYLSTVVTENHLYCEDNWKVPLTTSFFVGLGVSFGOLSDRFGRKNWLEATMAVQ 180
Db 121 kdfistvtenwlvnceddwtpltsffvgvlgsvsgqlsdrfgrkknwlfatmavq 180
Qy 181 TGFSLQLQFISISWEMFTVLVIVGMGQISNVYVAFILGTEILGKSVRIIFSTGLGVCTFFA 240
Db 181 tgfsvqlfistwemftvlfaivgmqgisnyvvafiglteilsksvriifstglgvctffa 240
Qy 241 VGYMLPLPAYFIRDRWMLLALTVPGLVCLVPLWFWFIPESPRWLISQRREAEEDIIOKA 300
Db 241 igymwplfayfirdwrmlaltlpglfcvplwfwfipesprwlisqrreaeedlioka 300
Qy 301 AKMNTAVPAVLDSVE--ELNPLKQOKAFILDLFRTNRNIAIMTMSLLMLTSLVGVYFA 358
Db 301 akmsvavpagifdplqelqelslkgqkvilldlfrtniatitvmavmlwmltsvgvfa 360
Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLKRTLPRTYIIAAVLFWGGGVLFIOIV 418
Db 361 lslvnpnlhgdvylncflslglievpayftawllrtlpriyiaagvlfwgggvllliqv 420
Qy 419 PDYVFLSLGLVLMGKFGITSAFSLMYVFTAEIPLTVRNMAVGVTSTASRVGSIAPYF 478
Db 421 pedyntvslglvmlgkfgitsafslmyvftaelypvlvrnmavgitmsarvgsiapyf 480
Qy 479 VYLGAYNRMLPYIVMGSLTVLIGIFTLFFPESLGMTLPETLEOMQVKVWFRSCKKTRDSM 538
Db 481 vylgaynrllpyilmgsltlvligitliffesgvtlpenleqmkgvrgfcgkktstvsv 540
Qy 539 ETEENPKVLITAF 551
Db 541 dreespkvlitaf 553

RESULT 4
AAY01650 AAY01650 standard; Protein: 557 AA.

XX AAY01650;

XX 23-JUN-1999 (first entry)

XX A protein with cation transporting activity.

XX Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
XX heart disease; cancer; anti-tumour drug; anticancer drug.

XX Homo sapiens.

XX WO9913072-A1.

XX 18-MAR-1999.

XX 07-SEP-1998; 98WO-JP04009.

XX 20-MAY-1998; 98JP-0156660.

XX 08-SEP-1997; 97JP-0260972.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Nezu J, Ose A;

XX WPI; 1999-215062/18.

XX N-PSDB; AAX26880.

XX Genes homologous with organic cation transporters OCT1 and OCT2,
XX useful in design of new drugs for treatment of diseases due to
XX abnormality of the transporter functions

XX Claim 1; Page 51-55; 97pp; Japanese.

XX The present sequence represents a protein with cation transporting
CC activity. The genes are significantly homologous with organic cation
CC

CC transporters OCT1 and OCT2. The genes may be used in drug development,
 CC particularly in the treatment of diseases due to abnormality of the
 CC organic cation transporter functions e.g. fatty liver, heart diseases
 CC and cancers, by controlling such as by inhibition or activation.
 CC Administration of anti-tumour and anticancer drugs in combination with
 CC a transporter protein inhibiting agent allows the agents to penetrate
 CC into the diseased cells to enhance the drug action.

XX Sequence 557 AA;
 Query Match 78.0%; Score 2218; DB 20; Length 557;
 Best Local Similarity 75.9%; Pred. No. 1.2e-217;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;
 Qy 1 MRDYDEVIAFLGEMGPQRLLIFLLSASIPNGFNGMSVFLAGTPEHRCRVPDAANLSS 60
 Db 1 mrdydeviaflgempqrllifllsasipngfngmsvflagtpchrcrvpdaanlss 60
 Qy 61 AWRNNSVPLRLDRGVPHSCSRYRLATIANFSALEPGRDVDLQLEQESCLDGEWFS 120
 Db 61 awrnhtvplrlrdgrevphscrryrlatiansaalgpegrdvdlqgleqescldgweifs 120
 Qy 121 QDVLSTVVTENLVCEDNKVPKLTSLFVGVLLGSFVSGQLSDRFGRKNVLFATMAVQ 180
 Db 121 qdvlstvtvtenlvcecdnkvpkltslfvgvllgslfvgvllgsfsgqlsdrgrknvlfatmgmq 180
 Qy 181 TGFSLQIFSIWEMFTVLFVIVGMQISNVVAVILGTEILGKSVRIIFSTLGVCTEFA 240
 Db 181 tgfslqifsknfemfvlvlgmgqisnyvaafvlgteilgksvriifstlglvcifya 240
 Qy 241 VGYMLLPFAFYFIDRMMLLALTVPGLVPLWTFIPESPRWLISQRFREAEDIIOKA 300
 Db 241 fgymlplfayfirdmmlaltvpglvplwtfipesprwlisqrfreaediioka 300
 Qy 301 AKMNTAVPAVIFD--SVEELNPKQKAFILDFTRTNIAIMTMSLLMLTSGVYFA 358
 Db 301 akangivvpstifdpseelqldskkqgshnildlftwlrmtmslmlwtisvgyfg 360
 Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIAAFLWGGVLLFIQIV 418
 Db 361 lsldapnlhgdaylncflsalieipayitawllrltpryiaaflwggvllfqiqlv 420
 Qy 419 PDVYFSLIGLVMLGKFGITSFASMLYVFTAEALYPTLVNRMAVGTSTASRVGSIAPYF 478
 Db 421 pddiylatvlvmvgkfgvtaafsmvvytaelyptlvnrmgvgsstasrfgsllspvf 480
 Qy 479 VYLGAYNRMLPYIVMGSILTVLIGITLFFPESLGMWLPETLEOMQKVKWFRSGK---KTR 535
 Db 481 vylgaydrfplylmgstliltaltlfpesfgtpltidqmlrvkmkhrktpshtr 540
 Qy 536 DSMETEENPKVL-ITAF 551
 Db 541 mlkdqgerptilkstaf 557

RESULT 5
 AA83929
 ID AA83929 standard; Protein: 557 AA.
 XX
 AC AA83929;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Human carnitine transporter protein OCTN2.
 XX
 KW Organic cation transportation; human; carnitine transporter protein;
 KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
 KW juvenile visceral steatosis.
 OS Homo sapiens.
 XX
 PN W0200014210-A1.

XX 16-MAR-2000.
 PD
 XX 07-SEP-1999; 99WO-JP04853.
 PF
 XX 07-SEP-1998; 98JP-0252683.
 PR
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA
 XX Nezu J, Oku A;
 PI
 XX WPI; 2000-256966/22.
 DR
 XX N-PSDB; AAA09889.
 DR
 XX Systemic carnitine deficiency gene OCTN2 encoding part of organic
 PT cation transporter, useful as diagnostic tool -
 PT
 XX Example 1; Page 36-40; 106pp; Japanese.
 PS
 XX This sequence represents the human carnitine transporter protein OCTN2.
 CC The coding sequence can be used as a target for diagnosis of systemic
 CC carnitine deficiency by detecting the presence of mutations in the
 CC sequence, especially seen in the disease juvenile visceral steatosis
 CC (jvs). The wild type OCTN2 gene can be used in the gene therapy of the
 CC disease state.
 CC
 XX Sequence 557 AA;
 SQ

Query Match 78.0%; Score 2218; DB 21; Length 557;
 Best Local Similarity 75.9%; Pred. No. 1.2e-217;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;
 Qy 1 MRDYDEVIAFLGEMGPQRLLIFLLSASIPNGFNGMSVFLAGTPEHRCRVPDAANLSS 60
 Db 1 mrdydeviaflgempqrllifllsasipngfngmsvflagtpchrcrvpdaanlss 60
 Qy 61 AWRNNSVPLRLDRGVPHSCSRYRLATIANFSALEPGRDVDLQLEQESCLDGEWFS 120
 Db 61 awrnhtvplrlrdgrevphscrryrlatiansaalgpegrdvdlqgleqescldgweifs 120
 Qy 121 QDVLSTVVTENLVCEDNKVPKLTSLFVGVLLGSFVSGQLSDRFGRKNVLFATMAVQ 180
 Db 121 qdvlstvtvtenlvcecdnkvpkltslfvgvllgslfvgvllgsfsgqlsdrgrknvlfatmgmq 180
 Qy 181 TGFSLQIFSIWEMFTVLFVIVGMQISNVVAVILGTEILGKSVRIIFSTLGVCTEFA 240
 Db 181 tgfslqifsknfemfvlvlgmgqisnyvaafvlgteilgksvriifstlglvcifya 240
 Qy 241 VGYMLLPFAFYFIDRMMLLALTVPGLVPLWTFIPESPRWLISQRFREAEDIIOKA 300
 Db 241 fgymlplfayfirdmmlaltvpglvplwtfipesprwlisqrfreaediioka 300
 Qy 301 AKMNTAVPAVIFD--SVEELNPKQKAFILDFTRTNIAIMTMSLLMLTSGVYFA 358
 Db 301 akangivvpstifdpseelqldskkqgshnildlftwlrmtmslmlwtisvgyfg 360
 Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIAAFLWGGVLLFIQIV 418
 Db 361 lsldapnlhgdaylncflsalieipayitawllrltpryiaaflwggvllfqiqlv 420
 Qy 419 PDVYFSLIGLVMLGKFGITSFASMLYVFTAEALYPTLVNRMAVGTSTASRVGSIAPYF 478
 Db 421 pddiylatvlvmvgkfgvtaafsmvvytaelyptlvnrmgvgsstasrfgsllspvf 480
 Qy 479 VYLGAYNRMLPYIVMGSILTVLIGITLFFPESLGMWLPETLEOMQKVKWFRSGK---KTR 535
 Db 481 vylgaydrfplylmgstliltaltlfpesfgtpltidqmlrvkmkhrktpshtr 540
 Qy 536 DSMETEENPKVL-ITAF 551
 Db 541 mlkdqgerptilkstaf 557

| | | | |
|--------|----------|---|-----|
| Db | 241 | fgfmwlpbfayfifrdwrmllallatvpvgvlgcalwvfipesprwlisggrikeaeaviirka | 300 |
| Qy | 301 | AKMNNITAVPAVIED--SVEELNPLKQOKAFILDLEFTRNIAIMTMSLLMLTSSVGYFA | 358 |
| Db | 301 | akngivapstifdpseqlqdnstkpqlhhiydliirtnirvitmsiilwltisvgyfg | 360 |
| Qy | 359 | LSIDAPNHLGDAYLNCFLSALIEIPAYITAWLLRTLPRLRYIIAAVLPWGGCVLLFTQLV | 418 |
| Db | 361 | lsldtpnlhgdinyncflaavevpayvialllqvlpryisaaiflgsgvllfmqlv | 420 |
| Qy | 419 | PVDYVFLSIGLWMLCKFGITSAFSMLKVFVTAELYPTLVNRNMAVGVTSTASRVGSIAPYF | 478 |
| Db | 421 | pelsfylvlstaivmvgkfgitsaysmyvytaelyptvtrnmvgvysstasrlgsispyf | 480 |
| Qy | 479 | VYLGAYNRMRLPYIVMGSTVVLIGITFLFPPEISLGMTLPETLEQMKQVKWFRSGK--- | 535 |
| Db | 481 | vyilgyadrflpyilmgslitaitltlfpesfgvpltdidgmrlvkgikqwdigsqr | 540 |
| Qy | 536 | DSMTEENPKVL-ITAF 551 | |
| Db | 541 | mqkdgeesptvikstaf 557 | |
| RESULT | 7 | | |
| ID | AAB20380 | | |
| ID | AAB20580 | standard; Protein; 557 AA. | |
| XX | AC | AAB20580; | |
| XX | AC | AAB20580; | |
| DT | DT | 11-DEC-2000 (first entry) | |
| XX | XX | Mouse OCTN2 amino acid sequence. | |
| DE | DE | Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter; | |
| XX | XX | identification; regulator; carnitine transport. | |
| OS | OS | Mus musculus. | |
| PN | PN | WO200046368-A1. | |
| PD | PD | 10-AUG-2000. | |
| PF | PF | 04-FEB-2000; 2000WO-JP00619. | |
| PR | PR | 05-FEB-1999; 99JP-0028406. | |
| PA | PA | (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC. | |
| PI | PI | Nezu J, Ose A; | |
| XX | XX | WPI; 2000-586982/55. | |
| DR | DR | Organic cation transporter gene OCTN3 expressed in testis for | |
| PT | PT | identification of regulators of carnitine transport for use as drugs | |
| XX | XX | Example 2; Fig 2; 58pp; Japanese. | |
| CC | CC | The present invention describes a mouse organic cation transporter | |
| CC | CC | protein (OCTN3). Also described are: (1) a method for screening | |
| CC | CC | compounds for their ability to regulate the transport of an organic | |
| CC | CC | cation into the cell, by generating a cell expressing OCTN3 at the | |
| CC | CC | cell membrane, contacting with the compound and organic cation, and | |
| CC | CC | observing the degree of transport of the organic cation; and (2) a | |
| CC | CC | method for screening compounds for their ability to be transported in | |
| CC | CC | the cell by OCTN3, by generating a cell expressing OCTN3 at the cell | |
| CC | CC | membrane, contacting with the compound and observing the degree of | |
| CC | CC | transport of the compound. OCTN3 can be used for the identification o | |
| CC | CC | regulators of the transport of organic cations (especially carnitine) | |
| CC | CC | into cells by OCTN3, for use as drugs. The present sequence represent | |
| CC | CC | the mouse OCTN2 amino acid sequence, which is used in an example from | |
| XX | XX | Sequence 557 AA; | |
| SQ | SQ | | |

DE Mouse OCTN3 protein SEQ ID NO:1.
XX Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
KW Identification; regulator; carnitine transport.
XX OS Mus musculus.
XX WO2000046368-A1.
XX 10-AUG-2000.
XX 04-FEB-2000; 2000WO-JP00619.
XX 05-FEB-1999; 99JP-0028406.
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Nezu J, Ose A;
XX WPI: 2000-586982/55.
XX N-PSDB; AAA88053.
XX Organic cation transporter gene OCTN3 expressed in testis for
PT Identification of regulators of carnitine transport for use as drugs
PT Claim 1; Page 34-39; 58pp; Japanese.
XX The present invention describes a mouse organic cation transporter
CC protein (OCTN3). Also described are: (1) a method for screening
CC compounds for their ability to regulate the transport of an organic
CC cation into the cell, by generating a cell expressing OCTN3 at the
CC cell membrane, contacting with the compound and organic cation, and
CC observing the degree of transport of the organic cation; and (2) a
CC method for screening compounds for their ability to be transported into
CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
CC membrane, contacting with the compound and observing the degree of
CC transport of the compound. OCTN3 can be used for the identification of
CC regulators of the transport of organic cations (especially carnitine)
CC into cells by OCTN3, for use as drugs. The present sequence represents
CC mouse OCTN3.
XX
SQ Sequence 564 AA;

Query Match 71.5%; Score 2034.5; DB 21; Length 564;
Best Local Similarity 69.0%; Pred. No. 6.5e-199;
Matches 380; Conservative 75; Mismatches 85; Indels 11; Gaps 3;

QY 1 MRDYDEVIAPLGEWGPFORLIRPLLSASIPNGFNMGMSVVELAGTPEHRCRVPDAANLSS 60
Db 1 mldydevtafigewgtfqrllfllsasiipngftgisavftaipchrcripdtvniss 60
QY 61 AWRNNSVPLRLRDRGVRPHSCSRYRLATIANFSALGLEPGRVDLQGLQESCLDGWEFS 120
Db 61 awrnhsipmetkdgpevpqkrcrryrlatianselglepgrdvdlqleqencldgweyd 120
QY 121 QDYLSTVTETNVLCDENKVKPLTSLFVGVLLGSFVSGQLSDRFGKKNVLFATMAVO 180
Db 121 kdiflstivtewdlvckddkaptlffvvgvllgsfsgqlsdrfgrknllfltmah 180
QY 181 TGFSELIQFISWEMTFLFVIVGMGOISNYVVAFTLGTETILGKSVRIITFSLGVCTFEA 240
Db 181 tgfslfqlfsvnfemflltylvgmghisnyvaafvlgtemisksvrllfatlgvciffa 240
QY 241 VGYMLPLFAFYIRDMRLMLLALTVPGLVCLPWLWFIPESPRWLISQRRFEARELIQKA 300
Db 241 fgfmvplfayfirewrrlllatlpgvlgcgalwffipesprwlisqgrikaeviirka 300
QY 301 AKMNTAVPAVIFDSVEELNPL-----KQKAFILDLFRTRNIAIMTMSLLMLWLTSV 354
Db 301 akingivapstifdp-setnklqddskkpkqshhdyldvrtptnrltlmsllwltisv 359
QY 355 GYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAWULLTLPLRYYIAAVLFWGGGVLLF 414

Db 360 gyfslsldtpnlgnyivncflllaavevpayvialwlllghvsrrysmagslfiggsvlll 419
QY 415 IOLVPDYDYFLSIGLVMGLKFGITSAFSLMYVFTFAELYPFLVRNMAVGVSTASRVGSII 474
Db 420 vqlvpsdlhlylstdtlvmvgkfgitsaysmvyvvytaelyptvvrnmvgvstasrllgsil 479
QY 475 APYEVYLGAYNRMLPYIVMGSLTVLIGLIFLFPPESSGMLPETLEOMQKVWFRSGKKT 534
Db 480 spfyvgyaydrripyllmsltliltaltiffessgvsipetidemqkvkl-----kg 535
QY 535 RDSMETEENPK 545
Db 536 rqlsllskkgspk 546

RESULT 10
AAB43038
ID AAB43038 standard; Protein; 560 AA.
XX
AC AAB43038;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human OREF2802 polypeptide sequence SEQ ID NO:5604.
XX
KW Human; open reading frame; OREFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO2000058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
XX N-PSDB; AAC77247.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4791-4792; 5507pp; English.
XX
CC AACT74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human OREFX open reading frames 1 to 3161. The OREFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORX-associated disorder. The
 CC nucleic acids can be used to express ORX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 560 AA;

Query Match 27.4%; Score 779.5; DB 21; Length 560;
 Best Local Similarity 32.9%; Pred. No. 7.7e-71;
 Matches 184; Conservative 113; Mismatches 214; Indels 49; Gaps 9;

QY 17 FORLFFLLSASIPNGFNGSVVFLAGTPEHRCRVP-----DAANLS 59
 Db 2 fqrviyfaicafniscgihyasvmgvtphhcrpgnsvqvfhnshnswlsdgtall 61
 QY 60 SAWRNNVPLRLRGREVPHS-CSRYRLATTANFSALGLEPRGVDVLGQLEQESCLDGE 118
 Db 62 ssqkdyvqlqngelwelsrcnrkr---entsslyge-----ytgskkefcvcdgyi 113
 QY 119 FSQDYVLTSTVTENLVCEENKVPVLTSLFFVGLVLSGVSGQSLDRGRKNVLPATMA 178
 Db 114 ydqntwkstavgtnlvcdkwlamlqlpmlfmgvllgsvtfyfsdrlgrvrlwatss 173
 QY 179 VQTGFSFLOISISWEMFTLVFVIVGMQISNVYVAIILGTEILGKSVRIIFSLGVCTP 238
 Db 174 smflfgiaaafavdyttfmaarfflamvasgylvgvfyvmefimgksr-twasvhlhsf 232
 QY 239 FAVGYMLPLFAYFTRDW---RMLLLATVPGVLCVPLWVNFIPESPRWLISORFREAE 295
 Db 233 favgtllvaltylvrtwlyqmlstvtvpfllcc---wvlpetpflwlsgryeaaq 289
 QY 296 IQQAANKMNTA-----VPAVIFDSVEELNPLKQKAFILDLFTRTNIAIMTMSLLW 349
 Db 290 lvdimakwnrasscklsellidlgppvsnsptevqkhnlsylfynwsikttrltvliw 349
 QY 350 MLTSGVFPALSDAPNLHGDAYLNCFLSALIEIPAYITAMLLLTLPRTYIIAIVLFWGG 409
 Db 350 ftgslgfyfslslnvgneylnflgvgveipaytfvciamdkvgrtrtvlayslfcsa 409
 QY 410 GVLLFIQVLPVDYFSLSIGLVMLGKFGTSAFSLMYVYTAELVPTLVRNMAVGVTSTASR 459
 Db 410 lacgvnmvlpkqhyilgvvtamvqkfaigaafglylytaelyptivrsiavsgsmvcr 459
 QY 470 VGSIIAPFVVLGYANRMLPYVINGMSTVLTIGITLFFPESLGMTLPETLEQMKQKWF 529
 Db 470 lasilapfsvdlsliwifpqlfvtgltmallsgvltklpeltgkrlattweeakl---- 525
 QY 530 SGKTRDSMETENPKVLIT 549
 Db 526 -----esenesksklllt 539

RESULT 11
 AAB49401
 ID AAB49401 standard; Protein; 540 AA.
 XX
 AC AAB49401;
 XX
 DT 02-MAR-2001 (first entry)
 XX
 DE Murine organic anion transporter 6.

XX Murine; organic anion transporter 6; mOATP6; cancer; inflammation;
 KW cardiovascular disease; central nervous system disorder; kidney disease;
 KW liver disease; autoimmune disease.
 XX Mus sp.
 OS WO200070048-A1.
 PN 23-NOV-2000.
 XX 15-MAY-2000; 2000WO-US13316.
 PF 14-MAY-1999; 99US-0134137.
 XX 12-MAY-2000; 2000US-0570293.
 PR (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX Feild J, Yue L, Ellens H;
 XX WPI; 2001-016235/02.
 DR N-PSDB; AAC83979.
 XX Murine organic anion transporter 6 polypeptide, useful for identifying
 PT agonists/antagonists that are useful in treatment of cancer, kidney
 PT disease, autoimmune disease, inflammation and cardiovascular disease
 XX Claim 2: Page 28-29; 32pp; English.
 PS The present sequence is murine organic anion transporter 6 (mOATP6).
 CC mOATP6 protein is useful for screening compounds which inhibit or
 CC stimulate the function of mOATP6 and also compounds that neither agonise
 CC nor antagonise OATP6. The identified agonists and antagonists are useful
 CC for prevention and treatment of human diseases, including cancer,
 CC inflammation, cardiovascular disease, central nervous system disorders,
 CC kidney diseases, liver disease and autoimmune diseases.
 XX Sequence 540 AA;

Query Match 24.5%; Score 698; DB 22; Length 540;
 Best Local Similarity 32.5%; Pred. No. 1.5e-62;
 Matches 183; Conservative 97; Mismatches 231; Indels 52; Gaps 19;

QY 4 YDEVIAFLGEMGPQ--RLIFFLLSASIPNGFNGSVVFLAGTPEHRCRVPDA-ANLS- 59
 Db 3 feelhkvvggfgpqlrnvlialprflpmhf--llpfmaavpahcalpdpapanish 60
 QY 60 -SAWRNNVPLRLRGREVPHS-CSRYRLATTANFSALGLEPRGVDVLGQLEQ-----SCL 114
 Db 61 qdlwkthlp-retgdsf--ssclrfaypqaipnvtlgtv---ynsgepegepltvpc 114
 QY 115 DQWFSQDYVLTSTVTENLVCEENKVPVLTSLFFVGLVLSGVSGQSLDRGRKNVLF 174
 Db 115 qgweydrsefsstiatetawdlvcqrglnkvtctcfllgavvyeylsdrfgrrill 174
 QY 175 ATMAVQTGSEFLOISISWEMFTLVFVIVGMQISNVYVAIILGTEILGKSVRIIFSLG 234
 Db 175 vavvstlalgmsaasvnyimfvtrmtlgsalagftiivlpelwldvhrtvagvis 234
 QY 235 VCTFFAVGYMLPLFAYFTRDWRMMLLATVPGVLCVP-----LWMFIPESPRWLISQRRF 290
 Db 235 t-tftwtggvllltlvgylrswrlllaatljp---cvpgllsiw-vpesarwlltqgrv 289
 QY 291 REAEDIQKAANKMNTAVPAVIFDSV--EELNPL-----KQKAFILDLFTRTNIAIMI 343
 Db 290 eakkyisicakngripse---dslsqeaalkvmtervsqirpsylldlrtsqirhsvl 346
 QY 344 MSLLMLMTSVGYFALSADPNLHGDAYLNCFLSALIEIPAYITAMLLLTLPRTYIIAIV 398
 Db 347 ccmnmwfgvnsfyygltidasglvtvqtlqllfavevpskitvflvgrlrlteag 406

QY 399 YIIAAVLFWGGVLLFTQLVFPVDYFELSICLVMLGKFGITSAFSMLYVFTAEIYPTLVN 458
DB 407 mlalaltfisi-----lvssdtkswitalvvgkafseaafttaylftselyptvlq 461
QY 459 MAVGVSTASRVGSIITAPYVYLGAYNRMPLPYIVMSLVLVIGIFTLFFPESLGMTLPET 518
DB 462 tgmftaligrigaslaplvllldgwwlllpklayggisflaactvlllpetkkaipet 521
QY 519 LEOMOKVKWFRSGKKTDRDSMETE 541
DB 522 iqdve-----rkgrki-drsgte 538
RESULT 12
AAW64538
ID AAW64538 standard; Protein; 554 AA.
XX
AC AAW64538;
XX
DT 21-OCT-1998 (first entry)
XX
DE Human liver cell clone HP01293 protein.
XX
KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
OS Homo sapiens.
XX
PN WO9821328-A2.
XX
PD 22-MAY-1998.
XX
PF 07-NOV-1997; 97WO-JP04056.
XX
PR 13-NOV-1996; 96JP-0301429.
XX
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX
DR WPI; 1998-297932/26.
DR N-PSDB; AAV49558, AAV49559.
XX
PT Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX
PS Claim 1; Page 96-98; 205pp; English.
XX
CC AAW64534-W64558 represent human proteins containing a transmembrane
CC domain. These proteins can be used for, e.g. research and nutrition, and
CC may have cytokine and cell proliferation/differentiation, immune
CC stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activity.
XX
SQ Sequence 554 AA;
Query Match 24.1%; Score 685.5; DB 19; Length 554;
Best Local Similarity 33.0%; Pred. No. 3e-61;
Matches 190; Conservative 87; Mismatches 229; Indels 69; Gaps 14;
QY 1 MRDYDEVIATFLGWGPFQR---LIFLLSASIIIPNGFMGVVFLAGTPEHRCRVPDAAN 57
DB 1 mptvddileqvgesgwfqkqafilcllsaaafap---icvgivfigftpdhchcspgva 57
QY 58 LSS--AWR-----NNSVPLRLDRDREVPVHSCRYR-----LATIA-NFSAIG 96
DB 58 lsqrcgspaeelnvtpvglgpageflgqcrreyevdnqsalscvdpplasiatnrshlp 117

QY 97 LEPRDNDLGLQEESCLDGHFESQDYVLSIVTVEWNLVCEDNWKPVLTTLSLFFVGVLLG 156
DB 118 lgp-----cqdgwvy--dtpgssivtefnlvocadswkldfscnagffg 162
QY 157 SFVSQOLSDRGRKNVLFATMAVQTGFSLQTFISISWEMFTVLEFVIVGMGQISNYVVAFI 216
DB 163 slgvgyfadrgrkclclgtvlvnavsgvlnafspnymsmllfrllgqlvskgnwmagyt 222
QY 217 LGTEILG----KSVRIEFTSLGVCTFFAVGVMLLPFLFAYFDRWRMLLALLATVPGVLCVP 272
DB 223 litetvsgsrrtvalmygma-----ftvgivaigtlayalphwrwqlavslptfll 277
QY 273 LWWFIPESPRWLISORREREADIIOKAKMNNTAVPAVI-----FDSVEELNPLKQOK 326
DB 278 ywcvpesprwllsqnrteaalkmdhiaqngklppadlkmleedvteklsp----- 332
QY 327 AFILDLFTRNTAINTMSLLMLTSSVGYFALSUDAENLHGDAYLNCFLSALIEIPAYI 386
DB 333 -sfadlfrtprlkrftfilmylftdsylvlqglilhmgatsgnyldflysalveipdaf 391
QY 387 TAWLLLRTPRRYIIAAVLFWGGVLLFTQLVFPVDYFELSICLVMLGKFGITSAFSMLYV 446
DB 392 ialitidrvgrlypmavsnllagaaclvmifispgdhwlnliimcvgrmgltciaqmici 451
QY 447 FTAELYPTLVNMAVGVTSTASRVGSIITAPYFVY-IGAYNRMPLPYIVMSLTVLIGIFTL 505
DB 452 vnaelyptfvnlgmvcslcdigiltpfivfrlrevwqalplllfavllgaagvtl 511
QY 506 FPEPSLGMTLPETLEOMQKVKWFRSGKKTDRDSMET 540
DB 512 llpetkgvalpetmkdaenl-----grkakpkent 541
RESULT 13
AAW51249
ID AAW51249 standard; Protein; 535 AA.
XX
AC AAW51249;
XX
DT 14-APR-2000 (first entry)
XX
DE Rat liver anion transporter protein OAT2.
XX
KW OAT2; anion transporter; liver; rat; dicarboxylic acid; prostaglandin;
KW non-steroid anti-inflammatory agent; anti-tumor.
XX
OS Rattus sp.
XX
PN JP11346779-A.
XX
PD 21-DEC-1999.
XX
PF 03-JUN-1998; 98JP-0169174.
XX
PR 03-JUN-1998; 98JP-0169174.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI; 2000-109694/10.
DR N-PSDB; AAZ44679.
XX
PT A liver-specific org. anion transporter - can transport various
PT substances including dicarboxylic acids
XX
PS Claim 2; Page 5-6; 13pp; Japanese.
XX
CC This invention describes a novel rat liver-specific org. anion
CC transporter OAT2. The liver-specific org. anion transporter OAT2
CC can transport various substances including dicarboxylic acids,
CC prostaglandins, non-steroid anti-inflammatory agents and anti-tumor
CC agents. This sequence represents the rat liver anion transporter OAT2
CC protein described in the method of the invention.
XX

```

SQ Sequence 535 AA;
Query Match 24.0%; Score 682; DB 21; Length 535;
Best Local Similarity 32.7%; Pred. No. 6.4e-61;
Matches 177; Conservative 97; Mismatches 230; Indels 38; Gaps 18;

QY 4 YDEVIATFLGEMGFQ--RLIFFLSASIIIPNGFNGMSVVFVLAGTPEHRCRPVDA-ANLS- 59
DB 3 fedllkvgqfgqfqlnvlmalpmlpmhf--llpyfmaavpahcalpapaalsh 60
QY 60 -SAWNNNSVPLRLDGRVPHSCSRYL-ATTANFSALGLEPGRVDVIGLEQE-----SC 113
DB 61 qdlwleahlp-retddsf--ssclrfayqtpvnt-igtevsns---gepegepltpvc 113
QY 114 LDGMEFQDQVYVSTVTEWNLNVEDNKKVPLTSLFFVGVILGSFVSGOLSDRFGRKNVL 173
DB 114 sqgweydrsefsatewdlvccqglnkfstctffgvlvgavvygylsdrfgrll 173
QY 174 FATMAVGTGFSFLQIPFSISWEMFTVLVIVGMGQISNVVAVPILGTEILGKSVRIIFSTL 233
DB 174 lvayvsslvlgmsaasinyimfvtrtitsalagftiivlplewldvehrtvagi 233
QY 234 GVCFFAVGYMLPLPFAFTRDRMRLLLALTVPGLVCP-----LWWEIPESPRMLISORR 289
DB 234 st-vfsgvgvlllalvgylirswillaatlp--cvpgiislw-vpesarwllcqr 288
QY 290 FREADIIQKAAMNNTAVPAVFDVSEELNPL-----KQKAFILDLFRTNRNIAITIM 344
DB 289 veeakyliscaklndrprvgesl-sqealnvvmeralgrpsyldlfrtsqrlrhisc 347
QY 345 SLLWMLTSGVYFALSIDAPNLHGDAYLNCFLSALIEIPAVITAMLLRLTPRYITAIV 404
DB 348 cmmwvfgvnyfsgvltldvsglgnvyqqlifgavelpskimvyflvrrlgrltaegm 407
QY 405 LFWGGGVLLP--TQLVPVDYFSLGVLMLGKFGTSAFMSLYVPTAEYPTLVNNAVG 462
DB 408 ll--gaaltfgtllvsletkswtalvvgvkafseaafttaylftselyptrvtrqtlg 465
QY 463 VTPASRVGSIAPFYFVYLGAVNRMPLPYVGMGLVLTIGIFTFPPESLGMLTLPETLEOM 522
DB 466 ltalmgrlgaslaplaalldgwvlllpkayvgialvaactallpetkkaqlpetiqdv 525
QY 523 QK 524
DB 526 er 527

RESULT 14
AAR77676
XX ID AAR77676 standard; Protein: 556 AA.
XX AC AAR77676;
XX DT 10-SEP-1996 (first entry)
XX DE Rat OCT-1 protein.
XX KW Rat; OCT-1; transporter protein; cationic; xenobiotic; pharmaceutical;
KW blood; liver; kidney; epithelial cell; intestine; tetraethylammonium;
KW proximal renal tubule cell; intestine; enterocyte; transgenic; renal;
KW biliary; excretion; resorption; modulator; uptake.
XX OS Rattus rattus.
XX FH Key Location/Qualifiers
XX FT Domain 20..46 "transmembrane domain"
XX FT Modified-site 71
XX FT Modified-site 97 "putative N-glycosylation site"
XX FT Modified-site 113 "putative N-glycosylation site"

```

```

FT FT Domain /note= "putative N-glycosylation site"
154..171
FT FT Domain /note= "transmembrane domain"
178..197
FT FT Domain /note= "transmembrane domain"
243..260
FT FT Domain /note= "transmembrane domain"
267..283
FT FT Domain /note= "transmembrane domain"
350..366
FT FT Domain /note= "transmembrane domain"
380..398
FT FT Domain /note= "transmembrane domain"
406..425
FT FT Domain /note= "transmembrane domain"
432
FT FT Modified-site
435..452 /note= "putative N-glycosylation site"
469..485 /note= "transmembrane domain"
494..514 /note= "transmembrane domain"
DE4424577-A1.
XX XX DE4424577-A1.
XX XX 18-JAN-1996.
XX XX 13-JUL-1994; 94DE-4424577.
XX XX 13-JUL-1994; 94DE-4424577.
XX XX (FARH ) HOECHST AG.
XX XX Gorboulev V, Gruendeman D, Koepsell H;
XX XX WPI; 1996-069422/08.
XX XX N-PSDB; AAT08702.
XX XX Transporter protein for cationic xenobiotic(s) and pharmaceuticals,
XX XX and related DNA and transformed cells - used e.g. to assess
XX XX excretion and resorption of cationic cpds.
XX XX Claim 1; Fig 2A1; 13pp; German.
XX XX This is the amino acid sequence of the rat OCT-1 protein, a new
XX XX transporter protein able to transport cationic xenobiotics and
XX XX pharmaceuticals from the blood into liver or kidney epithelial
XX XX cells or from the intestine. The gene was isolated by injecting
XX XX a rat kidney gene library into Xenopus laevis oocyte and isolating
XX XX clones conferring uptake of 14C-tetraethylammonium. One clone
XX XX designated OCT-1 was isolated. Expression of the gene was detect
XX XX in proximal renal tubule cells, in liver epithelial cell and in
XX XX intestinal enterocytes. The DNA can be used to generate transgenic
XX XX cells for use in in vitro test for renal/biliary excretion or
XX XX intestinal resorption of xenobiotics and pharmaceuticals. The protein
XX XX or cells expressing it can also be used to isolate modulators that
XX XX block uptake of pharmaceutical by the renal tubules.
XX XX Sequence 556 AA;
SQ

Query Match 24.0%; Score 681.5; DB 17; Length 556;
Best Local Similarity 33.0%; Pred. No. 7.6e-61;
Matches 183; Conservative 88; Mismatches 231; Indels 53; Gaps 11;

QY 1 MRDDEVIAFLGEMGFQRLIFELL---SASIIIPNGFNGMSVVFVLAGTPEHRCRPVDAAN 57
DB 1 mptvdvleqvgefgwqkqfllclisaslaf---iyygviyfgftpggyccpgvae 57
QY 58 LSS--AMR-----NNSVP-LRLRDGREVPHSCSRYL-----ATIANFSAL 95
DB 58 lsqrcgwsqaeehnytpvglgpsdeasflsqcmryevdwnqsgldcvplslvanrsq 117

```

Qy 96 GLEPGRDVLDGQLEQESCLDGEFSDVYVLTSTVTEWNLVCEENWVPLTSTLFFVGVLL 155
 Db 118 plp-----cehgwy--dtpgsivtefnlvcgdawkvdlfgscvnlqgffl 162
 Qy 156 GSFVQSLSRDRGRKNVLPATMAVOTGFSFIQIFSISWEMTVLFIIVGMOISNYVAF 215
 Db 163 gslvvyiadrgkicllvtlvtlsvsgvitavapdytsmlfrllqgmvskgswsgy 222
 Qy 216 ILGTEILGFSVR----IIFSTLGVCFTFAVGVMILLPLFAYFIRDRWMLLALVPGVLCV 271
 Db 223 tlitefvsgyrrttailqyma-----ftvlgvlagvayapdwrlqlavsiptflfi 277
 Qy 272 PLMWTFESPRWLSORRFEADIIQKAAKMNNTAVPAVIFDSVEELNPLKQOKAFILD 331
 Db 278 lyvfyvpsrwlslskrttravimeqlaqngkvppadlkmicleedasekrspsfad 337
 Qy 332 LFRTRNIAITMISLLMLTSGVGFALSLDAPNLHGDAYLNCFLSALIEIPAYITAWLL 391
 Db 338 lfrtptlrkhtvilmlyfscavlyggimhvgtatganlyldfysvivefpaafilvt 397
 Qy 392 LRTLPRYIIAAVLFWGGVLLFIQLVPDYVYFISIGLVMLGKFGITSAFSMLYVFTAEI 451
 Db 398 idrigtlyplaaenltvgaacillmifiphelhwnvtlactlgrmgativlqmvclvnael 457
 Qy 452 YPTLVNRMAVGTSTASRVGSIAPFYV-LGAYNRMLPYIVMGSLVLTIGITLFFPES 510
 Db 458 yptfirlngmvmcsalcldigiftbmfvrimevwqalpdlifglvltagatmlilpet 517
 Qy 511 LGMTLPETLEFQOKV 525
 Db 518 kqvalpetieaeai 532

RESULT 15
 AAB08823
 ID AAB08823 standard; Protein; 548 AA.
 AC AAB08823;
 XX 02-JAN-2001 (first entry)
 DE A human organic anion transporter 6 (OATP6).
 KW Human; organic anion transporter 6; OATP6; cancer; inflammation;
 KW cardiovascular disease; central nervous system disorder;
 KW autoimmune disease; kidney disease; liver disease.
 XX Homo sapiens.
 OS WO200052024-A1.
 XX 08-SEP-2000.
 PD 28-FEB-2000; 2000WO-US04952.
 PF 04-MAR-1999; 99US-0262563.
 PR (SMIK) SMITHKLINE BEECHAM CORP.
 PA Ellens H, Field J, Yue L;
 PI WPI: 2000-579270/54.
 DR N-PSDB; AAA75034.
 XX

OATP6 polypeptides and polynucleotides encoding them, useful for treating cancer, inflammation, cardiovascular disease, central nervous system disorders, autoimmune disease, kidney or liver disease -
 Claim 1; Page 31-32; 37pp; English.
 The present sequence represents human organic anion transporter 6 (OATP6) polypeptide. The OATP6 polypeptides and polynucleotides are useful for

CC treating cancer, inflammation, cardiovascular disease, central nervous
 CC system disorders, autoimmune disease, kidney and liver disease, or in
 CC identifying individuals who may handle drugs differently than normal
 CC individuals. They may also be used in identifying agonists and
 CC antagonists/inhibitors, which may be used in treating conditions
 CC associated with OATP6 imbalance, and to configure screening methods for
 CC detecting the effect of added compounds on the production of mRNA and
 CC polypeptides in cells. The polypeptides may be further used to
 CC establish assays to predict oral absorption and pharmacokinetics, and
 CC thus enhance compound and formulation design, as immunogens to produce
 CC immunospecific antibodies, and in identifying soluble or membrane-bound
 CC receptors. The polynucleotides may also be used as hybridisation probes
 CC or as primers for nucleic acid amplification, as diagnostic reagents,
 CC which may aid in the diagnosis of or susceptibility to a disease, and
 CC for chromosome identification.
 XX
 SQ Sequence 548 AA;

Query Match 23.0%; Score 653.5; DB 21; Length 548;
 Best Local Similarity 32.0%; Pred. No. 5.4e-58;
 Matches 180; Conservative 88; Mismatches 250; Indels 45; Gaps 18;
 Qy 4 YDEVIAFLGEGWGPQ--RLIFFLLSALIPNGFNGSVVFLAGTPEHRCRPDA-ANLS- 59
 Db 3 feeelqvggfgqlrnvallalprvlplhf--llpiflaavpahrcalpgapafsh 60
 Qy 60 -SAWRNNSVPLRLDRGREGVPHSCSYRLATIANFSAIGLEPGRDVDLQLEQE-----SCL 114
 Db 61 qdvwleahlp-repdg--tlssclrfayqpaltntllgee---rqsrgeledepatypcs 114
 Qy 115 DGWEFSQDVYLTSTVTE--WNLVCEDNWKVPLTTSLFFVGVLLGSFVSGQLSDRGRKNV 172
 Db 115 ggweydhsefstlasesqdlvceqqlnraastffagvlvgavagysdrfgrrfl 174
 Qy 173 LFATMAVQTGFSELIQFISWEMFTVLFIIVGNGQISNVYVAFILGTILGKSVRIIFST 232
 Db 175 llvayvstlvglaasaasvymfaiirtltsalagftlivmplelewidvehrvagv 234
 Qy 233 LGVCTTFVAGYMLLPFAYFIRDRWMLLALVPGVLCVP----LWTFIPSPRWLSQR 288
 Db 235 ls-stftwtggvmlalvgylirdwrlllavtlp---capgillslw-vpesarwiltqg 289
 Qy 289 RFREAEDIIQKAAKMNNTAVPAVIFDSVEELNPLKQOKAFI-----LDLFRNIAITMI 343
 Db 290 hvkeahryllhcarlngrpvcdsf-sgeavskvaagervrprsyldftrprlrhisl 348
 Qy 344 MSLLWMLTSGVGFALSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLPRYIIAA 403
 Db 349 ccvfwfvgvnyfysldvsglgnvyqqlfvgavelpskllvylsvryagriltqag 408
 Qy 404 VLPWGGGVLLF--IQLVFPDYVYFISIGLVMLGKFGITSAFSMLYVFTAEIYPTLVNRMAV 461
 Db 409 tll--gtalafgrtrllvssdmkswstvlavmgkafseaafttayiftselyptvlrqtgm 466
 Qy 462 GVTASTARVGSIIAPYFVYLGAYNRMLPYIVMGSLVLTIGITLFFPESLGMTLPETLEQ 521
 Db 467 gltalvgrlgslaplaallldgvmwslpkityggiallaagtalilpctrqaqpetiqd 526
 Qy 522 MQVKWFRSGKKTDRDSMETEENP 544
 Db 527 ver-----ksaptslqeemp 542

Search completed: August 16, 2001, 13:58:25
 Job time: 207 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:56:53 ; Search time 28.45 Seconds
(without alignments)
1475.296 Million cell updates/sec

Title: US-09-521-195-1

Perfect score: 2845
Sequence: 1 MRDDEVIAFLGEMGPQRL.....KKTRDSMETENPKVLITAF 551

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|----------|--------------------|
| 1 | 2218 | 78.0 | 557 | 2 JW0089 | organic cation tra |
| 2 | 2152 | 75.6 | 557 | 2 JE0346 | high-affinity carn |
| 3 | 754.5 | 26.5 | 576 | 2 T22509 | hypothetical prote |
| 4 | 699.5 | 24.6 | 593 | 2 JC4884 | organic cation tra |
| 5 | 681.5 | 24.0 | 556 | 2 S50862 | organic cation tra |
| 6 | 565 | 19.9 | 794 | 2 T27870 | hypothetical prote |
| 7 | 548 | 19.3 | 745 | 2 T16565 | hypothetical prote |
| 8 | 484.5 | 17.0 | 527 | 2 T01019 | transport protein |
| 9 | 475 | 16.7 | 539 | 2 C96758 | probable protein |
| 10 | 421.5 | 14.8 | 518 | 2 B86299 | hypothetical prote |
| 11 | 412.5 | 14.5 | 521 | 2 H86298 | hypothetical prote |
| 12 | 382.5 | 13.4 | 515 | 2 B96825 | hypothetical prote |
| 13 | 356 | 12.5 | 528 | 2 T21682 | hypothetical prote |
| 14 | 338 | 11.9 | 540 | 2 T25851 | hypothetical prote |
| 15 | 327 | 11.3 | 454 | 2 F75580 | hypothetical prote |
| 16 | 317 | 11.1 | 455 | 2 B83213 | probable sugar tra |
| 17 | 311 | 10.9 | 591 | 2 T30895 | sugar transport pr |
| 18 | 307 | 10.8 | 435 | 2 T15290 | hypothetical prote |
| 19 | 294.5 | 10.4 | 529 | 2 T23190 | hypothetical prote |
| 20 | 290 | 10.2 | 400 | 2 C65757 | transporter homolo |
| 21 | 286.5 | 10.1 | 478 | 2 T33985 | hypothetical prote |
| 22 | 281 | 9.9 | 422 | 2 G72234 | hypothetical prote |
| 23 | 276 | 9.7 | 461 | 2 D70073 | metabolite transpo |
| 24 | 275 | 9.7 | 520 | 2 T23545 | hypothetical prote |
| 25 | 274.5 | 9.6 | 443 | 2 E64725 | yaaU protein - Esc |
| 26 | 274.5 | 9.6 | 443 | 2 H85485 | probable transport |
| 27 | 272.5 | 9.6 | 442 | 2 A83122 | probable MFS trans |
| 28 | 271.5 | 9.5 | 524 | 2 T27082 | hypothetical prote |
| 29 | 271 | 9.5 | 423 | 2 T19030 | hypothetical prote |

ALIGNMENTS

RESULT 1

JW0089
organic cation transporter protein 2 - human

N:Alternate names: OCTN2

C:Species: Homo sapiens (man)

C:Date: 18-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 28-May-1999

C:Accession: JW0089

R:Wu, X.; Prasad, P.D.; Leibach, F.H.; Ganapathy, V.

Biochem. Biophys. Res. Commun. 246, 589-595, 1998

A:Title: cDNA sequence, transport function, and genomic organization of human OCTN2.

A:Reference number: JW0089; MUID:98289574

A:Accession: JW0089

A:Molecule type: mRNA

A:Residues: 1-557 <NUA>

A:Cross-references: GB:AF057164; NID:g3273740; PIDN:AAC24828.1; PID:g3273741

A:Experimental source: placenta

C:Comment: This transporter functions in the elimination of cationic drugs and other

Query Match 78.0%; Score 2218; DB 2; Length 557;

Best Local Similarity 75.9%; Pred. No. 1.6e-156;

Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

| | | | |
|----|-----|--|-----|
| Qy | 1 | MRDDEVIAFLGEMGPQRLIFELLSSAIIPIGNGFNGMSVVFVLAGTPEHRCRVPDAANLSS | 60 |
| Db | 1 | MRDYDEVTAFLGEMGPQRLIFELLSSAIIPIGNGFNGMSVVFVLAGTPEHRCRVPDAANLSS | 60 |
| Qy | 61 | AWRNSVPLRLRDGREGVPHSCSYRLATIANFSALGLEPGROVDLGQLESCLDGWEFS | 120 |
| Db | 61 | AWRNHTVPLRLRDGREGVPHSCRYRLATIANFSALGLEPGROVDLGQLESCLDGWEFS | 120 |
| Qy | 121 | QDVYLVTVTEWNLVCDNKNKPLTTSLFFVGVLLGSFVSGQLSDRFGKKNVLFATMAVQ | 180 |
| Db | 121 | QDVYLVTVTEWNLVCDNKNKPLTTSLFFVGVLLGSFVSGQLSDRFGKKNVLFATMAVQ | 180 |
| Qy | 181 | TGFSFLQIFSISWEMFTLVIVVGMQISNVVAVPILGTEILGKSVRIIFSTLGVCTFFA | 240 |
| Db | 181 | TGFSFLQIFSISWEMFTLVIVVGMQISNVVAVPILGTEILGKSVRIIFSTLGVCTFFA | 240 |
| Qy | 241 | VGYMLLPFAFIFDRWMLLALTVPGLVCLPWMEIPESPRWLIISQRRFRAEDIQKA | 300 |
| Db | 241 | FGYVVLPLFAFIFDRWMLLALTVPGLVCLPWMEIPESPRWLIISQRRFRAEDIQKA | 300 |
| Qy | 301 | AKMNTAVPAVIDF--SVEELNPLKQKQAFILDIFTRNIAITMTSLMLWMTSVGVFA | 358 |
| Db | 301 | AKANGIVVPTSTFIDPSLEQDLSKKQSHNILLDLRTWIRMTIMSLMWTISVGVFG | 360 |
| Qy | 359 | LSLDAPNLRGDAYLNCFLSALIEIPAYITAWLLRLPRYIIAAVLFWGGSVLLFIOLV | 418 |
| Db | 361 | LSLDTPLNLHGDIFVNCFLSAMVEVPAYLAWLLQLVLPRIYSKATALFGGSVLLFMQLV | 420 |
| Qy | 419 | PVDYFYSIGLVMLGKFGITSAFMSMLYVFTAEYLPTPLVRNMAVGVTSTASRVGSIAPYF | 478 |

probable sugar tra
hypothetical prote
hypothetical prote
hypothetical prote
hexose transport p
synaptic vesicle p
hypothetical prote
glucose transport
hexose transport p
glucose transport
probable transport
glucose transport
hypothetical prote
transmembrane tran
glucose transport
glucose transporte

```

Db 421 PPDLYLATVLMVWGKGVTAFAFNVVYTAELPTVVRNMGVGSSTASRLGSLSPYF 480
Qy 479 VYLGAYNRLPYIVMGSTVLIGITFTFFPSLGMTLPETLEOMOKVWFRSGK---KTR 535
Db 481 VYLGAYNRLPYIVMGSTVLIGITFTFFPSLGMTLPETLEOMOKVWFRSGK---KTR 540
Qy 536 DSMETEENPKVL-ITAF 551
Db 541 MLKDGQERPTLKSTAF 557

RESULT 2
JE0346
high-affinity carnitine transporter, CTL - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JE0346
R:Sekine, T.; Kusuhashi, H.; Utsumiya-Tate, N.; Tsuda, M.; Sugiyama, Y.; Kanai, Y.; Endo
Biochem. Biophys. Res. Commun. 251, 586-591, 1998
A:Title: Molecular cloning and characterization of high-affinity carnitine transporter
A:Reference number: JE0346; MUID:99011422
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-557 <SEK>
A:Cross-references: DDBJ:AB017260; NID:g3869208; PIDN:BAA34399.1; PID:g3869209

Query Match 75.6%; Score 2152; DB 2; Length 557;
Best Local Similarity 73.1%; Pred. No. 1.2e-151;
Matches 407; Conservative 66; Mismatches 76; Indels 6; Gaps 3;

Qy 1 MRDYDEVTAFLEMGPFQRLIFFLLSASIIPIGNGFMGSMVFLAGTPEHRCRVPDAANLSS 60
Db 1 MRDYDEVTAFLEMGPFQRLIFFLLSASIIPIGNGFMGSMVFLAGTPEHRCRVPDAANLSS 60
Qy 61 AWRNNSVPLRLDGRVPHSCSRRLATIANFSALGLEPGRDVLGQLESCLDGWEFS 120
Db 61 AWRNNSVPLRLDGRVPHSCSRRLATIANFSALGLEPGRDVLGQLESCLDGWEFS 120
Qy 121 QDYLSTVVTWENLVNEDNWKVPLTSLFFVGVLLGSFVSQGLSDRFGRKNVLFATMAVQ 180
Db 121 QDYLSTVVTWENLVNEDNWKVPLTSLFFVGVLLGSFVSQGLSDRFGRKNVLFATMAVQ 180
Qy 181 TGFSLQIFSISWEMFTVLVIVGVNGQISNVVAFILGTTELKSVRIIFSTLGVCTFFA 240
Db 181 TGFSLQIFSISWEMFTVLVIVGVNGQISNVVAFILGTTELKSVRIIFSTLGVCTFFA 240
Qy 241 VGYMLPLFAFYIRDRMLLALTPVGLCVPLWMTFIPESPRWLISQRRFREAEDIIQKA 300
Db 241 VGYMLPLFAFYIRDRMLLALTPVGLCVPLWMTFIPESPRWLISQRRFREAEDIIQKA 300
Qy 301 AKMNTAVPAVIFD---SYEELNPLKQKAFILDLFTRNIAIMTMSLLWMTSVGYPFA 358
Db 301 AKMNTAVPAVIFD---SYEELNPLKQKAFILDLFTRNIAIMTMSLLWMTSVGYPFA 358
Qy 359 LSIDAPNLHGDAYLNCFLSALIEIPAVITAWLLRLTPRYIIAAVLEWGGVGLFIQIV 418
Db 359 LSIDAPNLHGDAYLNCFLSALIEIPAVITAWLLRLTPRYIIAAVLEWGGVGLFIQIV 418
Qy 419 PVDYVFLSLGLVNLGKFGITSAFSLVYFAETPLVLRNMAVGVTSASRVGSIAPYF 478
Db 419 PVDYVFLSLGLVNLGKFGITSAFSLVYFAETPLVLRNMAVGVTSASRVGSIAPYF 478
Qy 479 VYLGAYNRLPYIVMGSTVLIGITFTFFPSLGMTLPETLEOMOKVWFRSGK---KTR 535
Db 479 VYLGAYNRLPYIVMGSTVLIGITFTFFPSLGMTLPETLEOMOKVWFRSGK---KTR 535
Qy 536 DSMETEENPKVL-ITAF 551
Db 541 MLKDGQERPTLKSTAF 557

```

```

RESULT 3
T22509
hypothetical protein F52F12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22509
R:Matthews, L.
submitted to the EMBL Data Library, December 1996
A:Reference number: Z19573
A:Accession: T22509
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-576 <WIL>
A:Cross-references: EMBL:Z83228; PIDN:CA805732.1; GSPDB:GN00019; CESP:F52F12.1
A:Experimental source: clone F52F12
C:Genetics:
A:Gene: CBSP:F52F12.1
A:Map position: 1
A:Introns: 16/1; 23/3; 50/1; 80/3; 108/2; 134/3; 221/1; 257/2; 319/3; 358/2; 490/3; 5

Query Match 26.5%; Score 754.5; DB 2; Length 576;
Best Local Similarity 30.2%; Pred. No. 2.8e-48;
Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;

Qy 3 DYDEVIAFLGEMGPFQRLIFFLLS-ASIIPIGNGFMGSMVFLAGTPEHRCRVPDAANLSSA 61
Db 3 DYDEVIAFLGEMGPFQRLIFFLLS-ASIIPIGNGFMGSMVFLAGTPEHRCRVPDAANLSSA 61
Qy 16 DDFVLEQVNGYGVYIVFFIICLTSLPSAFSAFNIPFVGNPHTCHIPESKEYLRP 75
Db 16 DDFVLEQVNGYGVYIVFFIICLTSLPSAFSAFNIPFVGNPHTCHIPESKEYLRP 75
Qy 62 WRNNSVPLRLDGRVPHSCSRRLATIANFSALGLEPGRDVLGQLESCLDGWEFSQ 121
Db 62 WRNNSVPLRLDGRVPHSCSRRLATIANFSALGLEPGRDVLGQLESCLDGWEFSQ 121
Qy 76 LTNDTOIL-----SCKQYNETQINVRFAFTAP-VDTYSDRISLVPCQNGWDYDN 124
Db 76 LTNDTOIL-----SCKQYNETQINVRFAFTAP-VDTYSDRISLVPCQNGWDYDN 124
Qy 122 DVLSTVVTWENLVNEDNWKVPLTSLFFVGVLLGSFVSQGLSDRFGRKNVLFATMAVQ 181
Db 122 DVLSTVVTWENLVNEDNWKVPLTSLFFVGVLLGSFVSQGLSDRFGRKNVLFATMAVQ 181
Qy 125 STYLDSELTVEFNLCQDQAWIEISTSFYVGSFGNCLFGYVADKFGRRSEFFVILT 184
Db 125 STYLDSELTVEFNLCQDQAWIEISTSFYVGSFGNCLFGYVADKFGRRSEFFVILT 184
Qy 182 GSFLOIFSISWEMFTVLVIVGVNGQISNVVAFILGTTELKSVRIIFSTLGVCTFFA 241
Db 182 GSFLOIFSISWEMFTVLVIVGVNGQISNVVAFILGTTELKSVRIIFSTLGVCTFFA 241
Qy 185 VCTASFAKDIESTIIRFFGTGLAFALQIPIIICMEFMGNSGR-IFSGMTSLFFGA 243
Db 185 VCTASFAKDIESTIIRFFGTGLAFALQIPIIICMEFMGNSGR-IFSGMTSLFFGA 243
Qy 242 GYMLPLFAFYIRDRMLLALTPVGLCVPLWMTFIPESPRWLISQRRFREAEDIIQKA 301
Db 242 GYMLPLFAFYIRDRMLLALTPVGLCVPLWMTFIPESPRWLISQRRFREAEDIIQKA 301
Qy 244 AMALLGVVAMFIRNRQLTFFCNAPFAFYIIVFFLPESPRWSVSGKWADAKKQLK 303
Db 244 AMALLGVVAMFIRNRQLTFFCNAPFAFYIIVFFLPESPRWSVSGKWADAKKQLK 303
Qy 302 KMN---NTAVPAVIFDSVEELNPLKQKAF-----ILDLFTRNIAIMTMSLLWMTS 353
Db 302 KMN---NTAVPAVIFDSVEELNPLKQKAF-----ILDLFTRNIAIMTMSLLWMTS 353
Qy 304 KMGKSNVDVDELV-DSMKNHONAAEKEKTRSHNVTDLFTKPNLRRKTLIVTYIWMNA 362
Db 304 KMGKSNVDVDELV-DSMKNHONAAEKEKTRSHNVTDLFTKPNLRRKTLIVTYIWMNA 362
Qy 354 VGYFALSADAPNLHGDAYLNCFLSALIEIPAVITAWLLRLTPRYIIAAVLEWGGVGL 413
Db 354 VGYFALSADAPNLHGDAYLNCFLSALIEIPAVITAWLLRLTPRYIIAAVLEWGGVGL 413
Qy 414 FTQLVPDVFYFSLGLVNLGKFGITSAFSLVYFAETPLVLRNMAVGVTSASRVGSI 473
Db 414 FTQLVPDVFYFSLGLVNLGKFGITSAFSLVYFAETPLVLRNMAVGVTSASRVGSI 473
Qy 423 SAMFMDDGVPWIVASASFTGKFGVSGGFAVIIFAGELPTVIVRAICMGSMSSVAGSGL 482
Db 423 SAMFMDDGVPWIVASASFTGKFGVSGGFAVIIFAGELPTVIVRAICMGSMSSVAGSGL 482
Qy 474 IAPYFVYLGAYNRLPYIVMGSTVLIGITFTFFPSLGMTLPETLEOMOKVWFRSGK 533
Db 474 IAPYFVYLGAYNRLPYIVMGSTVLIGITFTFFPSLGMTLPETLEOMOKVWFRSGK 533
Qy 483 LAPHIVNLGKIKVILLPLLMGLMALSAGILFFLPETLGLAPLPMPTIEDAENF-----GKK 537
Db 483 LAPHIVNLGKIKVILLPLLMGLMALSAGILFFLPETLGLAPLPMPTIEDAENF-----GKK 537
Qy 534 -----TRDSMETEENP 544
Db 534 -----TRDSMETEENP 544
Qy 538 PEPDSGMFTQAARKRESQP 556
Db 538 PEPDSGMFTQAARKRESQP 556

RESULT 4
JC4884
organic cation transporter protein 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
C:Accession: JC4884

```


C:Accession: T27870

R:White, S.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z20432

A:Accession: T27870

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-794 <WIL>

A:Cross-references: EMBL:Z66567; PIDN:CAA91492.2; GSPDB:GN00028; CESP:ZK455.8

A:Experimental source: clone ZK455

C:Genetics:

A:Gene: CESP:ZK455.8

A:Map position: X

A:Introns: 52/3; 119/2; 158/1; 207/3; 335/2; 428/3; 610/2; 636/3; 733/3

Query Match 19.9%; Score 565; DB 2; Length 794;

Best Local Similarity 26.4%; Pred. No. 4.2e-34;

Matches 158; Conservative 121; Mismatches 232; Indels 88; Gaps 20;

QY 3 DYDE---VIAFIGEWGPFQRLIFPLLSASIPNGFMGSMVFLAGTPEHRCRVP-----53

Db 155 DFEEDDLGILQIGCSYQWIIIVLIISVQVPHAMFNLSVVMYQPDHCKIPFNEE 214

QY 54 ----DAANLSSAW---RNSV---PURLRDGREVPH---SCSRYRLATIANFSALGLEPGRD 102

Db 215 SPSAELGYNTYTDQVLNSTIAFPRTFNKQRLHDDCHYFE-----RDYVHIKLSWAQ 270

QY 103 V-DLQLEQESCLDGEFSDQVYLSVTVTENLVCEENKVPYLTSLFFVGLGSRVSG 161

Db 271 VKDMATGKINCKEWEDTSMORTIIVTEWNVCDNNRAHVMSTSLGYLVGCFVGG 330

QY 162 QLSDRFGRKNVLFATMAYOTGFSFLQIFSISWEMFTVL-FVIVGQTSNTVVAIFLGT 220

Db 331 FISDRYGRKTAITGILTMFLGFLTYSKEFEIFLVVRELLAATNEAD-LAAVYLCME 389

QY 221 ILGKSVRIESTIGCTFFAVGYMLPLFAYFIRDMRLALLVPGLVPLWFPDES 280

Db 390 VTGTYRSIVGSL-IOAPWACQAFALAIAYLTYSWTMIHLICVLLLSIMLLYFLPES 448

QY 281 PRLISORFREAEIDIOAKANNNTAVPA--VIFDSVEELNPLK-OOKAFILDLFTRN 337

Db 449 PRLNLNKTQAKIIEACHYNNKSLPSDLGLVRAEKKWKHNEKPSYFHLPSSE 508

QY 338 IAIMTMSLLWN-----LTSVGY-----356

Db 509 LRFNRNVLFIVMWIOKNIFQNSRFQYSDNSGSLRNGCYIVRSVITRKVGRTLSKPYR 568

QY 357 -----FALSLOAFNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAAVLPWGGV 411

Db 569 NOKLCFRVRFD-----GNFFLNAMAGAEILPTLVFCVFLLR-MGRKRSQMLVLF-GSGL 621

QY 412 LFLIQLVPV--DYFLSLGLVMLKFGITSAFSMLYVFTAEYPTLVLRNMAVGTSTASR 469

Db 622 FLTTSVMVYRKQSTLALIEFLMLSKACIQGSFNILYIPTSELNPTVVRNSAVGISSMVAR 681

QY 470 VGSIIAPFVYLGNAR-MLPYIVMGSTVLIGITFLFFPESLGMTLPET-LEOMQVKV 526

Db 682 MGAGASGYAIIASLDVMTPLVPTMPTIFACFSLLAGCLVLLLPETQGLPLPDTLDSVQMVK 740

RESULT 7

T16565

hypothetical protein k05f1.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T16565

R:Wohlmann, P.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid K05F1.

A:Reference number: Z18537

C:Genetics:

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-745 <WOH>

A:Cross-references: EMBL:U29377; NID:g868173; PID:g868176; PIDN:AAA68713.1; CESP:K05F

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:K05F1.6

A:Introns: 70/2; 160/3; 205/3; 249/3; 293/2; 350/1; 386/3; 462/3; 546/2; 669/3; 680/3

Query Match 19.3%; Score 548; DB 2; Length 745;

Best Local Similarity 27.8%; Pred. No. 7e-33;

Matches 166; Conservative 112; Mismatches 233; Indels 86; Gaps 21;

QY 3 DYDEVIAFLGEMWPFQRLIFPLLSASIPNGFMGSMVFLAGTPEHRCRVP-----53

Db 112 DFGILNIIGGCRWQIWIIVLIALQOIPHAMFNLVVYMYDPEFCQMVPGENDTNDIT 171

QY 54 -----DAANLSSAWRN-NSVPLRLRDGREVPHSC-----SRYRLATIANFS 93

Db 172 ISSHGPYLMGVDDIKNISFVFPNANS-----DCAQYRDSCYFYERSEERYQLRRMPLE 225

QY 94 ALGLEPCRDVLDLQLEQESCLDGEFSDQVYLSVTVTENLVCEENKVPYLTSLFF-VG 152

Db 226 TAYSEAWKDV----APKKKC-QAYHEKDVMTETIVTDENLVC-DSWFAKGAHAMFYSIG 279

QY 153 VLLGSGFVSGQLSDRFGRKNVLFATMAYOTGFSFLQIFSISWEMFTVL-FVIVGQTSNTV 212

Db 280 YLLGCVLGGTASDKIGRKPITIGFGILSSMLGVLPFNDYYPMFLIRLLLSAICNEAADL 339

QY 213 VAFILGTEILGKSVRIESTIGCTFFAVGYMLPLFAYFIRDMRLALLVPGLVCP 272

Db 340 AAYTLCMEITGTYRAMVGSW-LQAPWALGALLAIYLTYSWTMIHLICVLLLSIMLLYFLPES 398

QY 273 LWFIFESPRLISORFREAEIDIOAKAK-----MNNTA-----VPA---VIFDSVEE 318

Db 399 FICISPEPRLIMVQNRVSEAEVIRKACREPPFPNMCCTTKCGMLPSDLELVSHREK 458

QY 319 LNLKOOKAFILDLFTRNTAINTMSLLMLTYSVGFALS-----DAPN---LHGDA 370

Db 459 LNKNGKGIGFLDLFTMKELRYRTISVCFVFNATALVYVGLVMAISDQSAPGRTLTGYF 518

QY 371 YLNCFLSALIEIPA-YITAWLLRLTPRYIIAAVLPWGGVLLFQLVLPVD-YVFLSIG 428

Db 519 HLNGTAGAEIETPLFACVMMW--QLGRKKALMLTITSGLFIIVAMLSVSGHYMLALA 576

QY 429 LVMLKFGITSAFSLYVFTAEYPTLVLRNMAVGTSTASRVGSIIAPYFVYLGNARML 488

Db 577 FMYFGKIAGVQAFNLIYIFTSELYPTVVRNTAVGVTSVAVRGSGLSYIALLS--NISL 634

QY 489 PYIVMGSLTVLIGITFLF-----FPESLGMTLPETLEQ-----MOKVWFRSGKKT 534

Db 635 PIVPM----IIFAVFSLFAGMLVFLVLPETSEKPLPETLDDAINFLPETKQFKTIEST 687

RESULT 8

T01019

transport protein homolog YUP8H12R.2 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 22-Oct-1999

C:Accession: T01019

R:Theologis, A.; Vysotskaya, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K

Oefner, P.; Davis, R.W.

submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.

A:Reference number: Z14227

A:Accession: T01019

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-527 <THE>

A:Cross-references: EMBL:AC020986; NID:g2494106; PID:g3152576; GSPDB:GN00059; ATSP:YU

A:Experimental source: cultivar Columbia

C:Genetics:

A:Gene: ATSP:YUP8H12R.2

RESULT 9

C96758
probable protein transporter t18k17.11 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96758
R:neurologists, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; D
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.;
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96758
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-539 <STO>
A:Cross-references: GB:AF005173; NID:g6598860; PIDN:AAF18714.1; CSPDB:GN00141
C:Genetics:
A:Gene: t18k17.11
A:Map position: 1

Query Match 14.8%; Score 421.5; DB 2; Length 518;
Best Local Similarity 25.5%; Pred. No. 1e-23;
Matches 138; Conservative 90; Mismatches 233; Indels 81; Gaps 14;

Qy 270 CVPLWTFIPESPRWLISQRRFREAEIIOKAAKMNTAVPAVIFDSVELNP---LKOOK 326
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Db 249 SIFIYFALESRWLHLEGNKEAIEVLKRIISPANGYLESV-----SSRLRPKETLEQTS 304
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Qy 327 AF-ILDLF----RTRNIAITMTMSLLIWMMLTSGYGFFALSLDAPNLHGDAYINCFLSALIE 381
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Db 305 SYSIKDLFIKWAFRRVTLVMIIMEGLM-----SIYGVPLAYVDIKVINYMSEALNAMVE 360
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Qy 382 IPAYITAWLLRLTPRYITAAVLFWGG--GULLFIQLVPVDY-----YFLSIGLVMLG 433
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Db 361 LPTFVFTPIILEQFSRSSRVLYNCILGGASGVLFCFMSL--YGRTKIAFALELGSTFCA 417
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Qy 434 KFGITSAFSMLYVFTEALEPVLTVLRNMAGVGTASTASRGVSIIAPFYFYLICAYNRMLPYIVM 493
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Db 418 RIG-----FNLMALIYELFPCTVRNSATMMKLQALVVGGACCPITASLGRNVPSLSFAVF 473
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Qy 494 GSLTVLIGIFTLPFPEISGMTLPETLEQMOK 524
: : : ||| | : : || : : : : : : : : : : : : : : : : :
Db 474 GFAMSGLGLFALLPETKGLSLCDTMEEOEQ 504
: : : ||| | : : || : : : : : : : : : : : : : : : : :

RESULT 13
T21682
hypothetical protein F32H5.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21682
R:Steward, C.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19458
A:Accession: T21682
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-528 <WTL>
A:Cross-references: EMBL:Z81524; PIDN:CAB04250.1; GSPDB:GN00023; CESP:F32H5.4
A:Experimental source: clone F32H5
C:Genetics:
A:Gene: CESP:F32H5.4
A:Map position: 5
A:Introns: 63/1; 108/3; 148/3; 249/2; 408/1; 472/3

[illegible][illegible]

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75580
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.V.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: F75580
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-454 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; MID:96460670; PIDN:AAF12486.1; PID:9646078
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0271
A:Map position: 2
C:Superfamily: yaaU protein

Query Match 11.5%; Score 327; DB 2; Length 454;
Best Local Similarity 26.0%; Pred. No. 8.8e-17;
Matches 107; Conservative 81; Mismatches 166; Indels 58; Gaps 13;

QY 143 PLTTSIF---FVGLGSEFVSQISDRGKNVLFATMAVQTGFSFLOIF--SISWEM-- 195
DB 67 PAATMLLTATFAGMLFGAWFGYLDNRVGRKRSVELTTVALGVVFGLAGALAPTLLWLLVA 126
QY 196 -FTVLFVIVGMQ(SNYVAFILGTEILGKSVRIIFSTGLGCTFFFAVGYMLPLFYFI- 253
DB 127 RELTGAIGGTLPIVDYSMAAEFVPTAWRGFLVYLES-----FWAVGTVVVAALAWWYS 180
QY 254 -----RDWRMLLLALTVPGLCVPLWFWIPESPRWLISORRPREAEDIIQKAAKNNTA 307
DB 181 TAPAPAEGRMWLLGLAALPGVLGVGTARTGIPDSRSLARGEQAQAALQKVAQANGGT 240
QY 308 VPVAFIDSVEELNPLKQOKAFILDIFRTNIAIMTMSLLIWMLTSGYFALSIDAPNL- 366
DB 241 LPAAPL-----AHPQPPRVSPAQLFR-GVLARTPLLMVTWGLSLGYGIFSWLPSPFL 294
QY 367 -----HGDAYLNQPLSALIEIPAYTTAWLLRLTPRYIIAAVLEWGG-GVLIETQLVP 419
DB 295 RAQGLDLGAVYRSTLLALAQVPGYLLAAYLVEKIGRVRTLVGFLTLGAVGAYLFLLAHD 354
QY 420 VDYVFLSIGLVMLGKFGITSAFSLYVFTAEYPTLVNRNMAVGTSTASRVGSIPIFYV 479
DB 355 ANTVLLTSALL--SFALLGAWGSLYAYTPELFTPLRTTGMGLVSGVARLASVWSP--- 408
QY 480 YLGAYNRMLPYIVMGSLTVLIGITFLFPF-----ESLGMTLPETLE 520
DB 409 SIGA-----MLLTGNLTALTAVCFVLAALAAWGIQVETRGQALAEAE 454

Search completed: August 16, 2001, 13:59:27
Job time: 154 Sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:58:28 ; Search time 44.83 Seconds
(without alignments)
1626.145 Million cell updates/sec

Title: US-09-521-195-1
Perfect score: 2845
Sequence: 1 MRDDEVIAFLGEMGPORL.....KKTRDSMETEENPKVLITAF 551

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues 425026

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_16:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_unclassified:*
- 13: sp_vertebrate:*
- 14: sp_virus:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|-------|-------------|
| 1 | 2845 | 100.0 | 551 | 4 | O14546 |
| 2 | 2833 | 99.6 | 551 | 4 | Q9H015 |
| 3 | 2484 | 87.3 | 553 | 11 | Q9R141 |
| 4 | 2470 | 86.8 | 553 | 11 | Q9R141 |
| 5 | 2034.5 | 71.5 | 564 | 11 | Q9WTN6 |
| 6 | 857 | 30.1 | 548 | 5 | Q9VCA2 |
| 7 | 852 | 29.9 | 548 | 5 | O01384 |
| 8 | 819 | 28.8 | 567 | 5 | Q9VCA3 |
| 9 | 773.5 | 27.2 | 561 | 5 | Q9V6L6 |
| 10 | 754.5 | 26.5 | 568 | 5 | Q9U539 |
| 11 | 754.5 | 26.5 | 576 | 5 | O02270 |
| 12 | 707 | 24.9 | 554 | 6 | O02270 |
| 13 | 703 | 24.7 | 555 | 4 | O15244 |
| 14 | 700 | 24.6 | 554 | 6 | O02713 |
| 15 | 699.5 | 24.6 | 553 | 11 | P70485 |
| 16 | 696.5 | 24.5 | 593 | 11 | O70577 |
| 17 | 695.5 | 24.4 | 555 | 11 | Q9RW2 |
| 18 | 695 | 24.4 | 674 | 5 | Q9VIK2 |
| 19 | 692 | 24.3 | 556 | 11 | O08966 |

| | | | | | |
|----|-------|------|-----|----|--------|
| 20 | 691.5 | 24.3 | 593 | 11 | P97558 |
| 21 | 688.5 | 24.2 | 554 | 4 | O15395 |
| 22 | 688.5 | 24.2 | 554 | 4 | Q9NQD4 |
| 23 | 687.5 | 24.2 | 556 | 11 | Q63089 |
| 24 | 687 | 24.1 | 556 | 11 | Q9RIQ4 |
| 25 | 684.5 | 24.1 | 554 | 4 | O15245 |
| 26 | 673 | 23.7 | 535 | 11 | O63314 |
| 27 | 672 | 23.6 | 557 | 5 | Q9VEX8 |
| 28 | 671 | 23.6 | 556 | 4 | O75751 |
| 29 | 653.5 | 23.0 | 456 | 4 | O14567 |
| 30 | 653.5 | 23.0 | 548 | 4 | Q9Y694 |
| 31 | 652 | 22.9 | 539 | 4 | Q9H2W5 |
| 32 | 650 | 22.8 | 551 | 11 | Q9WTW5 |
| 33 | 648 | 22.8 | 551 | 4 | Q9Y226 |
| 34 | 640 | 22.5 | 551 | 11 | O88446 |
| 35 | 638 | 22.4 | 537 | 11 | O88909 |
| 36 | 637 | 22.4 | 536 | 11 | Q9RIU7 |
| 37 | 627 | 22.0 | 538 | 5 | Q9V6H5 |
| 38 | 618.5 | 21.7 | 562 | 13 | O57379 |
| 39 | 604.5 | 21.2 | 545 | 11 | O61185 |
| 40 | 602.5 | 21.2 | 551 | 11 | O35956 |
| 41 | 599 | 21.1 | 553 | 11 | O54778 |
| 42 | 592 | 20.8 | 762 | 5 | Q9VNW8 |
| 43 | 586 | 20.6 | 542 | 5 | Q9VNX2 |
| 44 | 583.5 | 20.5 | 563 | 4 | O95742 |
| 45 | 583.5 | 20.5 | 569 | 5 | Q9VEY2 |

ALIGNMENTS

RESULT 1

O14546

ID O14546 PRELIMINARY; PRT; 551 AA.

AC O14546;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE POLYSPECIFIC ORANIC CATION TRANSPORTER.

GN OCTN1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCB1_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;

RX MEDLINE=98086199; PubMed=9426230;

RA Tamai I., Yabuuchi H., Nezu J., Sai Y., Oku A., Shimane M., Tsuji A.;

RT "Cloning and characterization of a novel human pH-dependent organic cation transporter, OCTN1.";

RL FEBS Lett. 419:107-111(1997).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.

DR EMBL; AB007448; BAA23356.1; -

DR InterPro: IPR001066; -

DR Pfam; PF00083; sugar_tr; 1.

DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.

KW Transmembrane.

SQ SEQUENCE 551 AA; 62176 MW; F5903421C789F60A CRC64;

Query Match 100.0%; Score 2845; DB 4; Length 551;

Best Local Similarity 100.0%; Pred. No. 1.3e-165;

Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDDEVIAFLGEMGPORLIFLLSASIIIPNGFMGMSVFLAGTPHRCRVPDANLSS 60

DB 1 MRDDEVIAFLGEMGPORLIFLLSASIIIPNGFMGMSVFLAGTPHRCRVPDANLSS 60

QY 61 AWRNNSVPLRLDGRVPHSCSRYLATIANFSALEPGRDVDLGOLESCLDGWEFS 120

DB 61 AWRNNSVPLRLDGRVPHSCSRYLATIANFSALEPGRDVDLGOLESCLDGWEFS 120

0x1997

```
QY 121 QDYVLTSTVTENLVCEENWVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
DB 121 QDYVLTSTVTENLVCEENWVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
QY 181 TGFSEFLQIFSISWEMFTVLFVIVGMQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
DB 181 TGFSEFLQIFSISWEMFTVLFVIVGMQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
QY 241 VGYMLLPFAFIRDRWMLLLALTVPGLVCPPLWPFIPESPRWLISQRRFREAEEDIIQKA 300
DB 241 VGYMLLPFAFIRDRWMLLLALTVPGLVCPPLWPFIPESPRWLISQRRFREAEEDIIQKA 300
QY 301 AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNIAIMTMSLLWMLTSVGYFALS 360
DB 301 AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNIAIMTMSLLWMLTSVGYFALS 360
QY 361 LDAPNLHGDAYLNCFLSALIEIPAYITAWLLRTLPRRYIIAAVLFWGGVLLFIQLVPV 420
DB 361 LDAPNLHGDAYLNCFLSALIEIPAYITAWLLRTLPRRYIIAAVLFWGGVLLFIQLVPV 420
QY 421 DYFSLISGLVLMGKFGITSAFSMLYVFTAEYPTLVNRNAVGVSTASRVGSIAPFYV 480
DB 421 DYFSLISGLVLMGKFGITSAFSMLYVFTAEYPTLVNRNAVGVSTASRVGSIAPFYV 480
QY 481 LGAYNRMLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMKVKWFRSGKTRDSMET 540
DB 481 LGAYNRMLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMKVKWFRSGKTRDSMET 540
QY 541 EENPKVLITAF 551
DB 541 EENPKVLITAF 551

RESULT 2
Q9H015
ID Q9H015 PRELIMINARY; PRT; 551 AA.
AC Q9H015:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE INTEGRAL MEMBRANE TRANSPORT PROTEIN.
GN UT2H.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Spritzberger F., Gruendemann D., Schoemig E.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y09881; CAA71007.1; -.
SQ SEQUENCE 551 AA; 62155 MW; C827A99AA78C9443 CRC64;

Query Match 99.6%; Score 2833; DB 4; Length 551;
Best Local Similarity 99.6%; Pred. No. 6.8e-165;
Matches 549; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRDYDEVIAFLGEGWGPQRLIFLLSASIIIPNGFNCGMSVFLAGTPEHRCRVPAANLSS 60
DB 1 MRDYDEVIAFLGEGWGPQRLIFLLSASIIIPNGFNCGMSVFLAGTPEHRCRVPAANLSS 60
QY 61 AWRNNSVPLRLRDGRVPHSCSRYRLATIANFSALEGPGRDVLGQLEQSCLDGWEEFS 120
DB 61 AWRNNSVPLRLRDGRVPHSCSRYRLATIANFSALEGPGRDVLGQLEQSCLDGWEEFS 120
QY 121 QDYVLTSTVTENLVCEENWVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
DB 121 QDYVLTSTVTENLVCEENWVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
QY 181 TGFSEFLQIFSISWEMFTVLFVIVGMQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
DB 181 TGFSEFLQIFSISWEMFTVLFVIVGMQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
```

```
DB 181 TGFSEFLQIFSISWEMFTVLFVIVGMQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
QY 241 VGYMLLPFAFIRDRWMLLLALTVPGLVCPPLWPFIPESPRWLISQRRFREAEEDIIQKA 300
DB 241 VGYMLLPFAFIRDRWMLLLALTVPGLVCPPLWPFIPESPRWLISQRRFREAEEDIIQKA 300
QY 301 AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNIAIMTMSLLWMLTSVGYFALS 360
DB 301 AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNIAIMTMSLLWMLTSVGYFALS 360
QY 361 LDAPNLHGDAYLNCFLSALIEIPAYITAWLLRTLPRRYIIAAVLFWGGVLLFIQLVPV 420
DB 361 LDAPNLHGDAYLNCFLSALIEIPAYITAWLLRTLPRRYIIAAVLFWGGVLLFIQLVPV 420
QY 421 DYFSLISGLVLMGKFGITSAFSMLYVFTAEYPTLVNRNAVGVSTASRVGSIAPFYV 480
DB 421 DYFSLISGLVLMGKFGITSAFSMLYVFTAEYPTLVNRNAVGVSTASRVGSIAPFYV 480
QY 481 LGAYNRMLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMKVKWFRSGKTRDSMET 540
DB 481 LGAYNRMLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMKVKWFRSGKTRDSMET 540
QY 541 EENPKVLITAF 551
DB 541 EENPKVLITAF 551

RESULT 3
Q9R141
ID Q9R141 PRELIMINARY; PRT; 553 AA.
AC Q9R141:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ORGANIC CATION TRANSPORTER OCTN1.
GN OCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu X., Wang H., Leibach F.H., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of OCTN1,
RT an organic cation transporter".
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF169831; AAD46922.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
SQ SEQUENCE 553 AA; 62362 MW; E26C8155768A14AD CRC64;

Query Match 87.3%; Score 2484; DB 11; Length 553;
Best Local Similarity 85.2%; Pred. No. 1.2e-143;
Matches 471; Conservative 36; Mismatches 44; Indels 2; Gaps 1;

QY 1 MRDYDEVIAFLGEGWGPQRLIFLLSASIIIPNGFNCGMSVFLAGTPEHRCRVPAANLSS 60
DB 1 MRDYDEVIAFLGEGWGPQRLIFLLSASIIIPNGFNCGMSVFLAGTPEHRCRVPAANLSS 60
QY 61 AWRNNSVPLRLRDGRVPHSCSRYRLATIANFSALEGPGRDVLGQLEQSCLDGWEEFS 120
DB 61 AWRNNSVPLRLRDGRVPHSCSRYRLATIANFSALEGPGRDVLGQLEQSCLDGWEEFS 120
QY 121 QDYVLTSTVTENLVCEENWVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
DB 121 QDYVLTSTVTENLVCEENWVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
QY 181 TGFSEFLQIFSISWEMFTVLFVIVGMQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
DB 181 TGFSEFLQIFSISWEMFTVLFVIVGMQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
```



```
GN OCr-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99227113; PubMed=10209228;
RA Wu X., Fei Y.J., Huang W., Chancy C., Leibach F.H., Ganapathy V.;
RT "Identity of the F52F12.1 gene product in Caenorhabditis elegans as an
RT organic cation transporter.";
RL Biochim. Biophys. Acta 1418:239-244(1999).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF110415; AAF21932.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 568 AA; 63505 MW; C0A3E73851F44056 CRC64;

Query Match 26.5%; Score 754.5; DB 5; Length 568;
Best Local Similarity 30.2%; Pred. No. 1.9e-38;
Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;

QY 3 DYDEVIAFLGEMGPQRILIFLLS-ASIIPNGFNGMSVVFAGTPEHRCRVPDANLSSA 61
DB 8 DDFDVLQGVNGYTIQIVFFFIICLPTSLPSAFSAFNPVFGNPPHTCHTPEGEKYL 67
QY 62 WRNNSVPLRLRDGREVPHSCSRYRLATIANFSALGLEPGRVDLGOLEQSCLDGWF 121
DB 68 LTNDTQIL-----SKQYNETQINVFRAFTSAP-VDTYSDRISLVPQNGWDYDN 116
QY 122 DYVLSTVVTWNLVCEDNWKVPLTTSFLFVGLGSGFVSGQLSDRFRGNKLVLFATMA 181
DB 117 STYLSLVTENFLVCDQAWIEISTTSFYVGSFGICNCLFGYVADKFGRRRSFFV 176
QY 182 GFSFLQIFSIWEMTVLFIIVGMCQISNVVAFILGTEILGKSVRIIFSLGVCVTF 241
DB 177 VCGTASSFAKDIESFIILRFTFTGLAFFALQIPFIICMEFGMNSGR-IFSG 235
QY 242 GYMLLPFAFYFIRDMRLMLLALTVPGLVCLPWLWFIPESPRWLISORRFAEDI 301
DB 236 AMALGVAMFIRRRQLTFFCNAPFAFYIIYFFLPSPRWSVSGKWADAKKQLK 295
QY 302 KMN---NTAVPAVIFDSVEELNPLKQKAF-----IDLFRTRNTAIMT 353
DB 296 KMNGKSNVDVDELV-DSMKNHQNAAEKETKRSNHNVDLFTPNLRRKTLIV 354
QY 354 VGYFALSADPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYITAAVLFW 413
DB 355 IYNGLTNLVSNLPVDDYWSFIINGAVELPGYFVVMVPLQLQACAGRWTLA 414
QY 414 FIQLVPVDYFYFLSIGLVMLKFGITSAFSLMYVFTAEIYPTLVRNMAVGVT 473
DB 415 SAMFMDGYPWLVASAFSGKGVSGFAVIYIFAGELYPTVVRIGMSSWAGSGL 474
QY 474 IAPYFVYLGAYNRMLPYIVMGSVLVLIGITFLFFPESIGMTLPETLEOMQKV 533
DB 475 LAPHIVNLGKIVKILPLLINGLMALSAGILTFELPETLGAPLPMTIEDAENP 529
QY 534 -----TRDSMETEENP 544
DB 530 PEPDSGMFTQAACKRESQP 548

RESULT 11
ID O02270 PRELIMINARY; PRT; 576 AA.
AC O02270;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
```

```
DE F52F12.1 PROTEIN.
GN F52F12.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX Matthews L.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighthouse J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; Z83228; CAB05732.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 576 AA; 64493 MW; F494EE94A7EBC0B1 CRC64;

Query Match 26.5%; Score 754.5; DB 5; Length 576;
Best Local Similarity 30.2%; Pred. No. 1.9e-38;
Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;

QY 3 DYDEVIAFLGEMGPQRILIFLLS-ASIIPNGFNGMSVVFAGTPEHRCRVPDANLSSA 61
DB 16 DDFDVLQGVNGYTIQIVFFFIICLPTSLPSAFSAFNPVFGNPPHTCHTPEGEKYL 75
QY 62 WRNNSVPLRLRDGREVPHSCSRYRLATIANFSALGLEPGRVDLGOLEQSCLDGWF 121
DB 76 LTNDTQIL-----SKQYNETQINVFRAFTSAP-VDTYSDRISLVPQNGWDYDN 124
QY 122 DYVLSTVVTWNLVCEDNWKVPLTTSFLFVGLGSGFVSGQLSDRFRGNKLVLFATMA 181
DB 125 STYLSLVTENFLVCDQAWIEISTTSFYVGSFGICNCLFGYVADKFGRRRSFFV 184
QY 182 GFSFLQIFSIWEMTVLFIIVGMCQISNVVAFILGTEILGKSVRIIFSLGVCVTF 241
DB 185 VCGTASSFAKDIESFIILRFTFTGLAFFALQIPFIICMEFGMNSGR-IFSG 243
QY 242 GYMLLPFAFYFIRDMRLMLLALTVPGLVCLPWLWFIPESPRWLISORRFAEDI 301
DB 244 AMALGVAMFIRRRQLTFFCNAPFAFYIIYFFLPSPRWSVSGKWADAKKQLK 303
QY 302 KMN---NTAVPAVIFDSVEELNPLKQKAF-----IDLFRTRNTAIMT 353
DB 304 KMNGKSNVDVDELV-DSMKNHQNAAEKETKRSNHNVDLFTPNLRRKTLIV 362
QY 354 VGYFALSADPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYITAAVLFW 413
DB 363 IYNGLTNLVSNLPVDDYWSFIINGAVELPGYFVVMVPLQLQACAGRWTLA 422
QY 414 FIQLVPVDYFYFLSIGLVMLKFGITSAFSLMYVFTAEIYPTLVRNMAVGVT 473
DB 423 SAMFMDGYPWLVASAFSGKGVSGFAVIYIFAGELYPTVVRIGMSSWAGSGL 482
QY 474 IAPYFVYLGAYNRMLPYIVMGSVLVLIGITFLFFPESIGMTLPETLEOMQKV 533
DB 474 IAPYFVYLGAYNRMLPYIVMGSVLVLIGITFLFFPESIGMTLPETLEOMQKV 533
```


Db 288 RWLISQNKIVKAMKIIKHIAKNGKSVPSLQNLTPDEDAGKKLKPISILDVVRTPQIRKH 347
Qy 342 TIMSLLWMLTSVGFALSLDAPNLHGD-AYLNCFLSALIEIPAYITAWLLRLTLPRRYI 400
Db 348 TLILMYNFTSSVLYXQGLIHHM-GLAGDNIYLDFFYSALVEFFPAFIIITIDRVGRYP 406
Qy 401 IAAVLFWGGVLLFPQLVPVDDYFELSIGLYMLGKFGITSFMSLYVFTAELYPTLVNRMA 460
Db 407 WAVSNMVAGAACLASVFIPDDQLWKITIACLGRMGITMAYEMVCLVNAELYPTIYIRNLG 466
Qy 461 VGVSTASRVGSIITAPYEVY-LGAYNRMLPYIYMGSLTVLIGITLFFPESLGMTLPETL 519
Db 467 VLVCSMCDIGIITPFLVFLTDIWMFFLVFVAVGVAGALVLLLPETGKALPETI 526
Qy 520 EQMKVKWFRSGKTR 535
Db 527 EDAENMQ--RPRKKER 540

Search completed: August 16, 2001, 14:03:35
Job time: 307 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:58:58 ; Search time 18.15 Seconds
(without alignments)
1039.932 Million cell updates/sec

Title: US-09-521-195-1

Perfect score: 2845

Sequence: 1 MRDYDEVIAFLGNGPFQRL.....KKTRDSMETENPKVLITAF 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 2218 | 78.0 | 557 | 1 | OCN2_HUMAN |
| 2 | 2165 | 76.1 | 557 | 1 | OCN2_RAT |
| 3 | 2164 | 76.1 | 557 | 1 | OCN2_MOUSE |
| 4 | 393 | 13.8 | 751 | 1 | YLX5_CAEEL |
| 5 | 317.5 | 11.2 | 1222 | 1 | YMP3_CAEEL |
| 6 | 307 | 10.8 | 435 | 1 | YT13_CAEEL |
| 7 | 294.5 | 10.4 | 529 | 1 | YOL1_CAEEL |
| 8 | 290 | 10.2 | 400 | 1 | YCE1_BACSU |
| 9 | 274.5 | 9.6 | 443 | 1 | YAAU_ECOLI |
| 10 | 270 | 9.5 | 461 | 1 | CSBC_BACSU |
| 11 | 262 | 9.2 | 445 | 1 | YGCS_ECOLI |
| 12 | 248 | 8.7 | 490 | 1 | GTR1_CHICK |
| 13 | 247 | 8.7 | 592 | 1 | HXT5_YEAST |
| 14 | 245 | 8.6 | 459 | 1 | YDJK_ECOLI |
| 15 | 245 | 8.6 | 566 | 1 | HKT2_KLULA |
| 16 | 244.5 | 8.6 | 567 | 1 | HXT9_YEAST |
| 17 | 243 | 8.5 | 451 | 1 | GTR1_PIG |
| 18 | 243 | 8.5 | 492 | 1 | GTR1_RAT |
| 19 | 240 | 8.4 | 495 | 1 | GTR3_CANFA |
| 20 | 238 | 8.4 | 492 | 1 | GTR1_BOVIN |
| 21 | 238 | 8.4 | 492 | 1 | GTR1_MOUSE |
| 22 | 237.5 | 8.3 | 452 | 1 | YDJE_ECOLI |
| 23 | 237 | 8.3 | 492 | 1 | GTR1_HUMAN |
| 24 | 237 | 8.3 | 546 | 1 | HXT0_YEAST |
| 25 | 237 | 8.3 | 567 | 1 | HXTA_YEAST |
| 26 | 236 | 8.3 | 451 | 1 | YJAJ_BACSU |
| 27 | 234.5 | 8.2 | 494 | 1 | GTR3_SHEEP |
| 28 | 234 | 8.2 | 491 | 1 | XYLE_ECOLI |
| 29 | 231 | 8.1 | 522 | 1 | GTR2_RAT |
| 30 | 226 | 7.9 | 457 | 1 | YIR0_YEAST |
| 31 | 225 | 7.9 | 492 | 1 | GTR1_RABIT |
| 32 | 225 | 7.9 | 763 | 1 | RGF2_YEAST |
| 33 | 223 | 7.8 | 413 | 1 | MUCK_ACICA |

| | | | | | |
|----|-------|-----|-----|---|------------|
| 34 | 222 | 7.8 | 523 | 1 | GTR2_MOUSE |
| 35 | 222 | 7.8 | 742 | 1 | SYV2_RAT |
| 36 | 221.5 | 7.8 | 570 | 1 | HXT6_YEAST |
| 37 | 221.5 | 7.8 | 570 | 1 | HXT7_YEAST |
| 38 | 220 | 7.7 | 496 | 1 | GTR3_HUMAN |
| 39 | 219.5 | 7.7 | 472 | 1 | ARAE_KLEOX |
| 40 | 218 | 7.7 | 569 | 1 | HXT8_YEAST |
| 41 | 215.5 | 7.6 | 472 | 1 | ARAE_ECOLI |
| 42 | 215.5 | 7.6 | 551 | 1 | HGT1_KLULA |
| 43 | 215 | 7.6 | 448 | 1 | PCAK_PSEPU |
| 44 | 215 | 7.6 | 457 | 1 | PCAK_ACICA |
| 45 | 215 | 7.6 | 493 | 1 | GTR3_MOUSE |

ALIGNMENTS

| | |
|------------|---|
| RESULT 1 | |
| OCN2_HUMAN | |
| ID | OCN2_HUMAN |
| AC | 076082; STANDARD; PRT; 557 AA. |
| DT | 01-OCT-2000 (Rel. 40, Created) |
| DT | 01-OCT-2000 (Rel. 40, Last sequence update) |
| DT | 01-OCT-2000 (Rel. 40, Last annotation update) |
| DE | ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22, MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER). |
| DE | SLC22A5 OR OCTN2. |
| OS | Homo sapiens (Human). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| OX | NCBI_TaxID=9606; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=98289574; PubMed=9618255; |
| RA | Wu X., Prasad P.D., Leibach F.H., Ganapathy V.; |
| RT | "cDNA sequence, transport function, and genomic organization of human OCTN2, a new member of the organic cation transporter family."; |
| RL | Biochem. Biophys. Res. Commun. 246:589-595(1998). |
| RN | [2] |
| RP | SEQUENCE FROM N.A. |
| RC | TISSUE=Kidney; |
| RX | MEDLINE=98352077; PubMed=9685390; |
| RA | Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M., Sai Y., Tsuji A.; |
| RT | "Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2."; |
| RL | J. Biol. Chem. 273:20378-20382(1998). |
| RN | [3] |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=99113835; PubMed=9916797; |
| RA | Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N., Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuishi T., Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M., Tsuji A.; |
| RT | "Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter."; |
| RL | Nat. Genet. 21:91-94(1999). |
| RN | [4] |
| RP | CHARACTERIZATION. |
| RX | MEDLINE=99384224; PubMed=10454528; |
| RA | Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H., Chen J., Conway S.J., Ganapathy V.; |
| RT | "Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter."; |
| RL | J. Pharmacol. Exp. Ther. 290:1482-1492(1999). |
| RN | [5] |
| RP | VARIANT CDS GLN-169. |
| RX | MEDLINE=99355597; PubMed=10425211; |
| RA | Burwinkel B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K., Gerbitz K.-D., Killmann M.W.; |
| RT | "Carnitine transporter OCTN2 mutations in systemic primary carnitine deficiency: a novel Arg169Gln mutation and a recurrent Arg282Ter |

mutation associated with an unconventional splicing abnormality.";
 Biochem. Biophys. Res. Commun. 261:484-487(1999).
 [6]
 RN VARIANT CDSP CYS-211.
 RX MEDLINE=99408248; PubMed=10480371;
 RA Vaz F.M., Scholte H.R., Rutter J., Hussaarts-Odijk L.M.,
 RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,
 RA Wanders R.J.A.;
 RT "Identification of two novel mutations in OCTN2 of three patients with
 RT systemic carnitine deficiency.";
 RL Hum. Genet. 105:157-161(1999).
 [7]
 RN VARIANT CDSP LEU-478.
 RX MEDLINE=99172075; PubMed=10072434;
 RA Tang N.L., Ganapathy V., Wu X., Seth P., Yuen P.M.,
 RA Wanders R.J., Fok T.F., Hjeltn N.M.;
 RT "Mutations of OCTN2, an organic cation/carnitine transporter, lead to
 RT deficient cellular carnitine uptake in primary carnitine deficiency.";
 RL Hum. Mol. Genet. 8:655-660(1999).
 [8]
 RN CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.
 RX MEDLINE=20026865; PubMed=10559218;
 RA Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;
 RT "Mutations in novel organic cation transporter (OCTN2), an organic
 RT cation/carnitine transporter, with differential effects on the
 RT organic cation transport function and the carnitine transport
 RT function.";
 RL J. Biol. Chem. 274:33388-33392(1999).
 [9]
 RN VARIANTS CDSP ARG-283 AND PHE-446.
 RX MEDLINE=20081068; PubMed=10612840;
 RA Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,
 RA Tsuji A.;
 RT "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a
 RT patient with primary systemic carnitine deficiency.";
 RL Hum. Mutat. 15:118-118(2000).
 [10]
 RN VARIANT CDSP LYS-452.
 RX MEDLINE=20145665; PubMed=10679939;
 RA Wang Y., Kelly M.A., Cowan T.M., Longo N.;
 RT "A missense mutation in the OCTN2 gene associated with residual
 RT carnitine transport activity.";
 RL Hum. Mutat. 15:238-245(2000).
 CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
 CC CARNITINE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,
 CC HEART AND PLACENTA.
 CC -1- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY
 CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE
 CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE
 CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND
 CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL
 CC MYOPATHY OR CARDIOMYOPATHY.
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
 CC CATION SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF057164; AAC24828.1; -;
 DR EMBL: AB015050; BAA29023.1; -;
 DR EMBL: AB016625; BAA36712.1; -;
 DR MIM: 603377; -;
 DR MIM: 212140; -;
 DR InterPro: IPR001066; -;

DR InterPro: IPR001687; -;
 DR Pfam: PF00083; sugar_tr; 1.
 DR PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
 KW Transport; Transmembrane; Glycoprotein; Disease mutation.
 FT TRANSMEM 21 163
 FT TRANSMEM 143 163
 FT TRANSMEM 173 193
 FT TRANSMEM 196 216
 FT TRANSMEM 233 253
 FT TRANSMEM 258 278
 FT TRANSMEM 343 363
 FT TRANSMEM 372 392
 FT TRANSMEM 414 434
 FT TRANSMEM 437 457
 FT TRANSMEM 489 509
 FT CARBOHYD 57 57
 FT CARBOHYD 64 64
 FT CARBOHYD 91 91
 FT VARIANT 169 169
 FT VARIANT 211 211
 FT VARIANT 283 283
 FT VARIANT 446 446
 FT VARIANT 452 452
 FT VARIANT 478 478
 FT MUTAGEN 352 352
 FT SEQUENCE 557 AA; 62751 MW; 928B1F6EFFG3C48D CRC64;

Query Match 78.0%; Score 2218; DB 1; Length 557;
 Best Local Similarity 75.9%; Pred. No. 4.2e-141;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;
 Qy 1 MRDYEVIATFGEWGPQRLIFFLLSIIIPNGFMGMSVVFVLAGTPHRCRVPDAANLSS 60
 Db 1 MRDYEVIATFGEWGPQRLIFFLLSIIIPNGFMGMSVVFVLAGTPHRCRVPDAANLSS 60
 Qy 61 AWRNNSVPLRLRDGRVPHSCRYRLATIANFSALEPGRDVDLQLEQESCLDGWEFS 120
 Db 61 AWRNHTVPLRLRDGRVPHSCRYRLATIANFSALEPGRDVDLQLEQESCLDGWEFS 120
 Qy 121 QDVYLSVTWTEWNLVCEENKVKPLTTSIFVGVLLGSFVSQSLSDRGKRVLFATMAVQ 180
 Db 121 QDVYLSVTWTEWNLVCEENKVKPLTTSIFVGVLLGSFVSQSLSDRGKRVLFATMAVQ 180
 Qy 181 TGFSLQIFTSWEMFTVFLVIVVGQISNVVAFILGTEILGSKSVRIIFSTLGVCTFFA 240
 Db 181 TGFSLQIFSKNFEMFVFLVVGQISNVVAFVILGTEILGSKSVRIIFSTLGVCTFFA 240
 Qy 241 VGYMLLPFAFIRDRWMLLALTPGVLCVPLWMTFPESPRMLISORREAEIIOKA 300
 Db 241 FGMYVLPFAFIRDRWMLLALTPGVLCVPLWMTFPESPRMLISORREAEIIOKA 300
 Qy 301 AKMNTAVPAVIFD--SVEELNPLKQKQKAFITLDRTRNIAIMTISLLWMLTISVGYFA 358
 Db 301 AKANGIVVPSTIFDPESELDLSSKKQKSHNLDLRTWNRWMTIMSILWMTISVGYFG 360
 Qy 359 LSLDAPNLHGDAYNCLFSLALIEIPAYITAWLLRLTPRRIYITAAVLFWGGVLLFLQIV 418
 Db 361 LSLDTPNLHGDIYVNCFLSANVEVPAYVLAWLLQYLPRIYSMAATALFLGSGVLLFNQIV 420
 Qy 419 PVDYFELSIGVLMGKFGITSAFSLMVYVFAEYPTLVLRNMAVGTSTASRVGSIIAPYF 478
 Db 421 PPDLYLATVLMVGKFGVTAFAFSWVYVYTAELPTVVRNMGVSVSTASRLSILSYF 480

QY 479 VYLGAYNRLPYIVMGSLTVLIGITLFFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535
DB 481 VYLGAYDRPLPYILMGSLTILTAITLFLPESFGTPTLTDQMLRVKMGKHKRTPSHTR 540
QY 536 DSMETENPKVL-ITAF 551
DB 541 MLKDGQERTILKSTAF 557
RESULT 2
OCN2_RAT STANDARD; PRT; 557 AA.
AC O70594; O90WLO;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER)
DE (USTR2R) (CTL).
GN SLC22A5 OR OCTN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=98200080; PubMed=9541011;
RA Schoenig E., Spitzberger F., Engelhardt M., Martel F., Oerding N.,
RA Gruendemann D.;
RT "Molecular cloning and characterization of two novel transport
RT proteins from rat kidney";
RL FEBS Lett. 425:79-86(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;
RX MEDLINE=99011422; PubMed=9792817;
RA Sekine T., Kusuha H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,
RA Kanai Y., Endou H.;
RT "Molecular cloning and characterization of high-affinity carnitine
RT transporter from rat intestine";
RL Biochem. Biophys. Res. Commun. 251:586-591(1998).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
RA Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
RT cation transporter 2 (OCTN2), an organic cation/carnitine
RT transporter";
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CC CARNITINE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES
CC AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND
CC ARTERIOLES IN THE HEART, IN THE LABYRINTHINE LAYER OF THE
CC PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE
CC BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CC CATION SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ001933; CAA05106.1; --

DR EMBL; AB017260; BAA34399.1; --
DR EMBL; AF110416; AAD54059.1; --
DR InterPro; IPR001066; --
DR Pfam; PF00083; sugar_tr; 1.
KW PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
KW Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 21 41
FT TRANSMEM 143 163
FT TRANSMEM 173 193
FT TRANSMEM 198 218
FT TRANSMEM 233 253
FT TRANSMEM 258 278
FT TRANSMEM 342 362
FT TRANSMEM 374 394
FT TRANSMEM 407 427
FT TRANSMEM 431 451
FT TRANSMEM 489 509
FT CARBOHYD 57 57
FT CARBOHYD 64 64
FT CARBOHYD 91 91
FT CONFLICT 275 275 W > G. (IN REF. 2).
SQ SEQUENCE 557 AA; 62567 MW; 172472E7B0A5F043 CRC64;
Query Match 76.1%; Score 2165; DB 1; Length 557;
Best Local Similarity 73.2%; Pred. No. 1.5e-137;
Matches 408; Conservative 68; Mismatches 75; Indels 6; Gaps 3;
QY 1 MRDYDEVIATLGEWGPQRLFFLFFLSASIPNGFNGSVVFLAGTPEHRCRVPDAAALSS 60
DB 1 MRDYDEVIATLGEWGPQRLFFLFFLSASIPNGFNGSVVFLAGTPEHRCRVPDAAALSS 60
QY 61 AWRNSVPLRLRDGVRPHSCSRRLATIANFSALEPGRDVDLQLEQESCLDGEFSS 120
DB 61 AWRNSHPIETKDGQVQPCRRYRLATIANFSALEPGRDVDLQLEQENCLDGEYN 120
QY 121 ODVYLSTVVTWNLVCEENKVPILTSLFFVGVLLGSFVSGQLSDRGKRVLPATMAVO 180
DB 121 KDVFLSTIVTWDLVCKDDWKAPLITSLFFVGVLLGSFVSGQLSDRGKRVLPATMAVO 180
QY 181 TGFSLQFISISWEMFTVLFVIVMGQISNYVAVFILGTILGKSVRIIFSTLGVCTFFA 240
DB 181 TGFSLQFISVNFEMFTVLFVIVMGQISNYVAVFILGTILGKSVRIIFATLGVCTFFA 240
QY 241 VGYMLLPFAFIRDRWMLLALTPVGVLPVLPWFIPESPRWLISQRRPREADITKA 300
DB 241 FGVVLPFAFIRDRWMLLALTPVGVLPVLPWFIPESPRWLISQRRPREADITKA 300
QY 301 AKMNTAVPAVFD--SVEELNPLKQKAFITLDFTRTNIAIMTMSLLMLLSVGYFA 358
DB 301 AKFNIVAPSTIFDSELDLNSKKPQSHHYYDLVTRNRIRIIMSLMLLTSVGYFG 360
QY 359 LSLDAPNLHGDAYLNCFLSALIEPAYITAWLLLTLPRTYIIAAVLPWGGVLLFTQLV 418
DB 361 LSLDAPNLHGDIYVNCFLAALAAVEPAVLAALLQLHLPRTYISAAFLGGSVLLFTQLV 420
QY 419 PVDYFLSGLVLMKCFKQITSAFSLMYVFTAEIPTLVLRNMAVGTSTASRGVSIAPYF 478
DB 421 PSELVFTALVWVGKFGITSAFSLMYVFTAEIPTLVLRNMAVGTSTASRGVSIAPYF 480
QY 479 VYLGAYNRLPYIVMGSLTVLIGITLFFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535
DB 481 VYLGAYDRPLPYILMGSLTILTAITLFLPESFGTPTLTDQMLRVKMGKHKRTPSHTR 540
QY 536 DSMETENPKVL-ITAF 551
DB 541 TOKDGESFTVLKSTAF 557
RESULT 3
OCN2_MOUSE
ID OCN2_MOUSE STANDARD; PRT; 557 AA.
AC Q920E8;

01-OCT-2000 (Rel. 40, Created)
01-OCT-2000 (Rel. 40, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).
GN SLC22A5 OR OCTN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99113835; PubMed=9916797;
RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,
Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuihi T.,
Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,
Tsujii A.,
RT "Primary systemic carnitine deficiency is caused by mutations in a
gene encoding sodium ion-dependent carnitine transporter.";
Nat. Genet. 21:91-94 (1999).
[2]
SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.
RC STRAIN=C3H;
RX MEDLINE=99057546; PubMed=9837751;
RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;
RT "A missense mutation of mouse OCTN2, a sodium-dependent carnitine
cotransporter, in the juvenile visceral steatosis mouse.";
RL Biochem. Biophys. Res. Commun. 252:590-594 (1998).
[3]
SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
cation transporter 2 (OCTN2), an organic cation/carnitine
transporter.";
RL J. Pharmacol. Exp. Ther. 290:1482-1492 (1999).
CC -!- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CARNITINE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL
STEATOSIS (JVS).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CATION SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL: AB015800; BA336590.1; -
DR EMBL: AF111425; AAC99787.1; -
DR EMBL: AF110417; AAD34060.1; -
DR InterPro: IPR001066; -
DR Pfam: PF00083; sugar.tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
KW Transport; Transmembrane; Glycoprotein; Disease mutation.
FT TRANSMEM 21 41
FT TRANSMEM 143 163
FT TRANSMEM 173 193
FT TRANSMEM 198 218
FT TRANSMEM 233 253
FT TRANSMEM 258 278
FT TRANSMEM 342 362
FT TRANSMEM 374 394
FT TRANSMEM 407 427
FT TRANSMEM 431 451
POTENTIAL.

FT TRANSMEM 489 509
FT CARBOHYD 57 57
FT CARBOHYD 64 64
FT CARBOHYD 91 91
FT CARBOHYD 322 322
FT VARIANT 352 352
SQ SEQUENCE 557 AA; 62779 MW; 6093F0EB9612B204 CRC64;
L -> R (IN JVS).
Query Match 76.1%; Score 2164; DB 1; Length 557;
Best Local Similarity 72.9%; Pred. No. 1.7e-137;
Matches 406; Conservative 70; Mismatches 75; Indels 6; Gaps 3;
QY 1 MRDYEVIATFGEWPPQRLIFFLLSIIIPNGFNGSVVFLAGTPEHRCRVDPDANLSS 60
DB 1 MRDYEVIATFGEWPPQRLIFFLLSIIIPNGFNGSVVFLAGTPEHRCRVDPDANLSS 60
QY 61 AWRNNSVPLRLDRGVRPHSCSYRLATIANFALSGLPGRDVLDGLQEQESCLDGWEFS 120
DB 61 AWRNHSIPLTKDGRQVPQKRRYRLATIANFSELGLEPGRDVLDGLQEQESCLDGWEYD 120
QY 121 ODVYLSVTVTENNLVCEENKVPVLTSLFFVGVLLGSFVSGQLSDRFGKKNVLFATMAVQ 180
DB 121 KDVFSLTIVTMDLVCKDQWKAPLTTSLFFVGVLLGSFVSGQLSDRFGKKNVLFATMAVQ 180
QY 181 TGFSLQIFSISSWEMFTVLFVIVGMGOISNYVAVFILGTILGKSVRIIFSTGLGVCTFFA 240
DB 181 TGFSLQVFSVNEFMFTVLFVIVGMGOISNYVAVFILGTILGKSVRIIFATLGVCIFFA 240
QY 241 VGYMLPLFAFYFIRDMRLMLLALTVPGLVLCVPLWFTIPSPRWLISORRFREAEIIQKA 300
DB 241 FGMVLPFAFYFIRDMRLMLLALTVPGLVLCVPLWFTIPSPRWLISORRFREAEIIQKA 300
QY 301 AKMNTAVPAVIDF--SVEELNPLKQKAPILDETRNIAIMTMSLLMLTSGVYFA 358
DB 301 AKINGIVAPSTIFDPSELQDLNSTKQHLHIYDILTRNIRVITIMSIILWLTISVGYFG 360
QY 359 LSLDAPNLHGDAYLNCFLSALIEIPATITAWLLRLTPRYITIAAVLFWGGVLLFIQLV 418
DB 361 LSLDTPNLHGDIYVNCFLLAAVEVPAYVLAWLLQVLPRLYSISAAFLGGSVLLFMQLV 420
QY 419 PVDYFELSLGVMLGKGIITSFASMLYVFTAEYLPVLRNMAVGVSTASRVGSIITAPYF 478
DB 421 PSELFIYSTALVMVGKGIITSAYSMYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 480
QY 479 VYLGAYNRLPYIVMGSLTVLIGITLFFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535
DB 481 VYLGAYDRELPIYILMGSLTITLITLFFPESFGVPLPTIDQMLRVKGIQWQIOSQTR 540
QY 536 DSMETENPKVL-ITAF 551
DB 541 MQKDGEESTVLKSTAF 557

RESULT 4
YLX5_CAEL STANDARD; PRT; 751 AA.
AC P46501;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOPHETICAL 84.8 KDA PROTEIN P23F12.5 IN CHROMOSOME III.
GN F23F12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U12965; AAA20607.1; -;
 DR WormPep: F23F12.5; CE01252.
 DR InterPro: IPR001066; -;
 DR Pfam: PF00083; sugar_tr; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 57 77 POTENTIAL.
 FT TRANSMEM 102 122 POTENTIAL.
 FT TRANSMEM 137 157 POTENTIAL.
 FT TRANSMEM 200 220 POTENTIAL.
 FT TRANSMEM 235 255 POTENTIAL.
 FT TRANSMEM 320 340 POTENTIAL.
 FT TRANSMEM 348 368 POTENTIAL.
 FT TRANSMEM 379 399 POTENTIAL.
 FT TRANSMEM 410 430 POTENTIAL.
 FT TRANSMEM 432 452 POTENTIAL.
 FT TRANSMEM 515 535 POTENTIAL.
 FT TRANSMEM 547 567 POTENTIAL.
 FT TRANSMEM 583 603 POTENTIAL.
 FT TRANSMEM 614 634 POTENTIAL.
 FT TRANSMEM 678 698 POTENTIAL.
 SQ SEQUENCE 751 AA; 84832 MW; A6C4F43540295EFC CRC64;

Query Match 13.8%; Score 393; DB 1; Length 751;
 Best Local Similarity 27.0%; Pred. No. 3.5e-19;
 Matches 119; Conservative 91; Mismatches 184; Indels 46; Gaps 14;
 QY 127 TVVTEWNLVCEDN-KWVPLTTSLEFVGLGSGVSDRGKRVLFATMAVQTGFSF 185
 DB 303 SWQDDKFCGKAYDAWVAWTAIQVILGAIYTGHLGDHFGKRPVSFFGSLGILGV 362
 QY 186 LQIFSIWSEMTVLFVIVGQISVNVVAFILGTILGKSVRIIFSTLGVCTFFAVGY-- 243
 DB 363 ASGFAPSWVEFAAFRIVGTSIASILIVFYAYILEFIEPEQRVELR-----SFFNNGYAR 417
 QY 244 MLLPLPAYFIRDMRLALLTVPGVLCVPLWFIPESPRWLISQRRFREAEDIIQAAKM 303
 DB 418 LVFTLACFCIGYWRSAAIATSLSLPLPVLVLLPESPKWFTKKRRDARAARVAVL 477
 QY 304 NNTAVPAV-----IFDSVEELNPLKQKAFIL-DLFRTRNIAIMTISLLWMLTSV 354
 DB 478 --SGIYVNDQDSIEISEKLEE-----KTKIYTKMDLFTSWIATYRTIVGSLWFSTSL 531
 QY 355 GYFALSDAPNLHGDAYLNCFLSALIEIPAYITAMLLLTLP---RRYI-----IAAVLF 406
 DB 532 SAFGSDLSNGLAGNLYLQFVSGAVTAFKIFVFLDITVPSFDRRLHQYQIAMLIC 591
 QY 407 WGGGVLLFTOLVPV-----DIYFLSIGLVMGKFGITSAFSLMYFTAEPLTVLR 457
 DB 592 Y--CVINVLMLPESDCGSGSRDLAIIINTIGVSFIBT--WDACYLVAVECEPPTKIR 647
 QY 458 NMAVGVTSTASRVGSIAPFVYVVLGAYNRMPLVYVNGSL-TVLIGITLTFPPESLGMTLP 516
 DB 648 TIGIGTCSLARTGALLAQMAVLSDIYRPAPYAVVCSIGTISLLISCVFLPDTKGVDL- 706
 QY 517 ETLEQMKVKWFRSGKKTRD 536
 DB 707 AALDPTTELDYDRKKSMTEN 726
 RESULT 5
 YMP3_CAEEL STANDARD; PRT; 1222 AA.
 ID YMP3_CAEEL

AC Q10947;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HYPOTHETICAL 139.9 KDA PROTEIN B0361.3 IN CHROMOSOME III.
 GN B0361.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=BRISTOL N2.
 RA Du Z.;
 RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U00031; AAA50622.1; -;
 DR WormPep: B0361.3; CE00752.
 DR Pfam: PF00083; sugar_tr; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 263 283 POTENTIAL.
 FT TRANSMEM 307 327 POTENTIAL.
 FT TRANSMEM 329 349 POTENTIAL.
 FT TRANSMEM 355 375 POTENTIAL.
 FT TRANSMEM 382 402 POTENTIAL.
 FT TRANSMEM 413 433 POTENTIAL.
 FT TRANSMEM 442 462 POTENTIAL.
 SQ SEQUENCE 1222 AA; 139868 MW; CBA42A80A254FB80 CRC64;
 Query Match 11.2%; Score 317.5; DB 1; Length 1222;
 Best Local Similarity 20.3%; Pred. No. 6.4e-14;
 Matches 114; Conservative 106; Mismatches 219; Indels 123; Gaps 14;
 QY 3 DYDEVIAFLGSEWGPFP-----QRLIFFLISAIIPNGFMGMSVVFAGTPEHRCRVPD 54
 DB 35 DPDRFVEAYGAYGKYQIFTVVLVQTLNFFYSSSMYI-----NSFVQL--NLEKQCE-- 83
 QY 55 AANLSSAWRNNSVPLRLDGRVPHSCSRVRLATIANFSALEPGRDVDLQLEQESCL 114
 DB 84 -----YKNETIP-----ISETCQ-----IETESKAFGNLNGEYC- 113
 QY 115 DGWEFSODVYL-----STVTEWNLVCEDNKMKPLTTSLEFVGLGSGVSDRGKRVLD 166
 DB 114 --GAENTLVNVTNQKASTNLIVDFDLSCSHWFFQFGLTIFTIGAVIAPVPMSLADR 170
 QY 167 FGRKNVLFATMAVQTGFSFQIFSIWSEMTVLFVIVGQISVNVVAFILGTILGKSV 226
 DB 171 YGRKPIIVTTAILAFLANMAASFSFNAIFLILRAFIACGSDSYLSVASVATCEYLSEKA 230
 QY 227 RIIFSTLGVCTFFAVGYMMLPLFAYFIRDMRMLLATVPCVLCVPLWFIPESPRWLIS 286
 DB 231 R-AWITVYVNVAVSLGMVMTLLVLTMTDDMRVYFVLSPLGVGFALWYFLPESPHMLIT 289
 QY 287 QRRFREAEIIQKAAKMNNTAVPAVIEDSVBELNPLKQKAFILDLFRTRNIAIMTISL 346
 DB 290 KNKTEKLYIKTANRM-----VISL 310
 QY 347 LLWMLTSVGYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAMLLLTLPRTYIAAVLF 406

```

Db 311 V-----YFAISFMSVLEGGDQVOAFLYSSLIEIPAGLAVIPLMMKMRKMIWICLV 362
QY 407 WGG-----GVLLFQLPVVDYIFLSIGLVMLGKFGITSAFSMLYVFTAEYPTLVLRNMAVG 462
Db 363 FQTLALIGVTVFL-----DSYEFKLVMLVAKVMATIIYSVHPINATEOFFPYSVSLCFS 417
QY 463 VTSASRVGSITAPYVYIGAYNMLPYIVMGSLTVLIGTIFLFFPESLGMTLPETLEQM 522
Db 418 LNMIPSGMGIISPYYVKHIVMSPNPFVFFVIALFSPISATLAFMLHETKNKKLPDIESL 477
QY 523 -----QKVKVPFSGKKTRDSM 538
Db 478 SYPSTNDLSAYRRSKSSSSV 499

RESULT 6
YT13_CAEEL
ID YT13_CAEEL STANDARD; PRT; 435 AA.
AC Q10917;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 48.6 KDA PROTEIN B0252.3 IN CHROMOSOME II.
GN B0252.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Du Z., Waterston R.;
RA Halloran N., Green P., Thierry-Mieg J., Olu L., Dear S., Coulson A.,
RA Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,
RA Ainscough R., Waterston R.;
RT "The C. elegans genome sequencing project: a beginning.";
RL Nature 356:37-41(1992).
RN [2]
RP REVISIONS.
RC STRAIN-BRISTOL N2;
RA Durbin R.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U23453; AAC46757.1; -
CC WormPep: B0252.3; CE02419.
CC InterPro: IPR001066; -
CC Pfam: PF00083; sugar_tr; 1.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 55 75 POTENTIAL.
CC FT TRANSMEM 384 404 POTENTIAL.
CC SEQUENCE 435 AA; 48571 MW; 5F6160359FA1840B CRC64;

Query Match 10.8%; Score 307; DB 1; Length 435;
Best Local Similarity 23.5%; Pred. No. 1.le-13;
Matches 94; Conservative 73; Mismatches 155; Indels 78; Gaps 8;

QY 128 VVTENLVCEDNMKVPLTSTLFFGVGLGVSFVSGQLSDRFGRKNVLPATMAVQGFSLQ 187
Db 91 VADEFDLTGDSALAESITTFYVGNMIGGMFIPPLADHYGLRPLVFFATVLLMAVGMIS 150
QY 188 IFSISWENFTLVFVVGQISNVVVAEIPLITELIGKSVRIIFTGLVCTFVAVGYMLLP 247
Db 151 AFSTSIMMFCIMRMIHGFYTAAGLAGVLYGENTPLRLR-FFTSVYFVGMVWVAGCFGL 209
QY 248 LFAYFIQWRMLLALTPGV-LCVPLWETPESPRMLISORREAREEDIQK-AARKNN 305
Db 210 LLAYILPDRVLMFCISVPNFVALLIYVTPESLHFLVSSQQNEKTEAWLEKIRGPKGD 269
QY 306 TAVPAVIFDSVEELNPLK-----QQKAFILDIFTRNIAIMTISLLWMLTSSVGYFALS 361
Db 270 ISASDIVERDENGSSFKTLICREIKTKFTTLFQ----- 302

```

```

QY 362 DAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAALVFLWGGVLLFIQLVPVD 421
Db 303 -----DRIYI-----VLFWIGIILLYEFGKPL- 324
QY 422 YFISIGLVMLGKFGITSAFSMLYVFTAEYPTLVLRNMAVGTVTASRVGSIIAPYFVYL 481
Db 325 FEPCA-----HEFGRSSSLHFFSDPHEQIFPTDGRKNKICGFCETLSRFGGMLSPYLSHL 378
QY 482 GAYNRMLPYIVMGSLTVLIGTIFLFFPESLGMTLPETLEQ 521
Db 379 TAVHALAPAITLSLIAVSGGLLTLLPETLNTKLPSTIAE 418

RESULT 7
YOUL_CAEEL
ID YOUL_CAEEL STANDARD; PRT; 529 AA.
AC P30636; Q21101;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III.
GN ZK637.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=92168156; PubMed=1538779;
RA Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,
RA Halloran N., Green P., Thierry-Mieg J., Olu L., Dear S., Coulson A.,
RA Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,
RA Ainscough R., Waterston R.;
RT "The C. elegans genome sequencing project: a beginning.";
RL Nature 356:37-41(1992).
RN [2]
RP REVISIONS.
RC STRAIN-BRISTOL N2;
RA Durbin R.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11115; CAA77460.1; -
CC EMBL: Z22175; CAA77460.1; JOINED.
CC EMBL: Z22175; CAA80131.1; -
CC EMBL: Z11115; CAA80131.1; JOINED.
CC PIR: S15786; S15786.
CC WormPep: ZK637.1; CE066638.
CC InterPro: IPR001066; -
CC Pfam: PF00083; sugar_tr; 1.
CC PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
CC PROSITE: PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
CC KW Hypothetical protein; Duplication; Transmembrane; Transport.
CC FT TRANSMEM 86 106 POTENTIAL.
CC FT TRANSMEM 122 142 POTENTIAL.
CC FT TRANSMEM 158 178 POTENTIAL.
CC FT TRANSMEM 238 258 POTENTIAL.
CC FT TRANSMEM 320 340 POTENTIAL.
CC FT TRANSMEM 373 393 POTENTIAL.
CC FT TRANSMEM 411 431 POTENTIAL.
CC FT TRANSMEM 482 502 POTENTIAL.
CC SEQUENCE 529 AA; 58317 MW; 8D2FF4CBA15ECD2D CRC64;

```



```
Query Match      10.4%; Score 294.5; DB 1; Length 529;
Best Local Similarity 25.6%; Pred. No. 9e-13;
Matches 110; Conservative 78; Mismatches 150; Indels 91; Gaps 20;

QY 134 LVCDNDKV-----PLTTSFFVGLVSGSVQSLSDRFQ-RKNVLFAF-----MAVQTG 182
DB 110 LACE-WGISVQQAALVTCVFSQMLSTFWCKICDRGRRKGLFSLVACIMGVISG 167
QY 183 FSFLQIFSISWEMFTVLF-----VIVMGQISNVVAF--ILGTEILGKSVRIIFSLGVC 236
DB 168 -----MSPHFYVLLFFRGLGFGIGGVPQSVTLVYAEFLPTAQRKACVLLIES----- 214
QY 237 TEFAVGVMLLPLPAYFIRD---WRMLLALTVP-GVLCVPLWVFIPESPRWLISQRRFRE 292
DB 215 -FWAIGAVFALLAYFVNESGFRMLFSLPLGIFAVASFW-LPESARFDMASGHPER 272
QY 293 AEDIIQKAAMNNTAVP-AVIFDSVEELNPLKQOKAFIL--DLFRTNIAIMFIMSLLLW 349
DB 273 ALETLQAAARNRVOLPTGRVSVSTRAGSESRGDIANLLSPDLRKT-----TILLWCW 326
QY 350 MLTSVGVFALSADPNL-----HG-----DAYLNCFLSALIEIPAY 385
DB 327 AITAFSYGMVLETTVLFQSHDECHGFLSNGTQMEVCQPLTRSDYFDLLSTTLAEFPGL 386
QY 386 ITAWLLL-----RTLPRRYIIAAVLEFWGGVLLFIQLVPVDYFYFLSIGVLMGKFGITS 439
DB 387 IITVLLIIEWFRKRTMALEYAVFA-----IFTELLYFCLDRFTVTV-LIFVARAFISG 438
QY 440 AFSMLYVFTAEYPTLVRNMAVGVTSTASRVGSIIAPYFYVILGAYNMLPYIVMGSLITVL 499
DB 439 AFQCAVYVTEVPTTLRAVGLGTCSEMARIGAIVA-----SEKSLSPIGIGYTAAIL 492
QY 500 IGIPTLFFP 508
DB 493 GLIASLSIP 501

RESULT 8
YCEI_BACSU
ID YCEI_BACSU STANDARD; PRT; 400 AA.
AC O34691;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN RAPJ-OPUAA INTERGENIC
DE REGION.
GN YCEI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kumano M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000617; BAA22256.1; -
DR EMBL; Z99105; CAB12089.1; -
```

```
Subtilist; BG12773; yceI.
InterPro; IPR001066; -.
Pfam; PF00083; sugar_tr; 1.
PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane.
FT TRANSMEM 15 35 POTENTIAL..
FT TRANSMEM 50 70 POTENTIAL..
FT TRANSMEM 78 98 POTENTIAL..
FT TRANSMEM 99 119 POTENTIAL..
FT TRANSMEM 143 163 POTENTIAL..
FT TRANSMEM 166 186 POTENTIAL..
FT TRANSMEM 218 238 POTENTIAL..
FT TRANSMEM 254 274 POTENTIAL..
FT TRANSMEM 281 301 POTENTIAL..
FT TRANSMEM 305 325 POTENTIAL..
FT TRANSMEM 344 364 POTENTIAL..
FT TRANSMEM 371 391 POTENTIAL..
SQ SEQUENCE 400 AA; 43708 MW; E0AE0CEE5DD27395 CRC64;

Query Match      10.2%; Score 290; DB 1; Length 400;
Best Local Similarity 25.5%; Pred. No. 1.3e-12;
Matches 105; Conservative 68; Mismatches 165; Indels 74; Gaps 16;

QY 116 GHEP-SQDV-YLSTVV-----TENNLVCEQ-NWKVPLTTSLFVGVLLGSFVSGQLSDRFQ 168
DB 19 GLFDAMDVGILSFITAAALHVEWNLSPPEMKW-----IGSVNSIGMAGAFLLGLADRIQ 74
QY 169 RKNVLFATMAVOTGFSFLQIFSISWEMFTVLFVIVMGQISNVVAFILGTEIL-----G 223
DB 75 RKVFIITLLCSIGSGISAFVTSLSAFLILFRVIGMLGGLGELPVASTLVSEAVVPERG 134
QY 224 KSVRIIFSTGLVCTFFAVGYMLLPLFAYFI---RDWRMLLLALTVPVGLCVPLWVFIPE 280
DB 135 RVIVLLES-----FWAVGWLAALISYFVTPSGWQAALLLTALTAFYALVLRSLDPS 188
QY 281 PRWLISQRRFREAEEDIIQKAAMNNTAVPAVIFDSVEELNPLKQOKAFILDLFRTNIAI 340
DB 189 PKY-----ESLSAKRRSM-----WENKSV-----WARQYIR 215
QY 341 MTIMSLLLMLTSGVGFALSADPN---LHGDAYLNCF---LSALIEIPAYITAWLLR 393
DB 216 PTVMLSIVWFCVFSYGMFLWLPSPVLLKGFESMIQSEYVLLMTLAQLPGYFSAWLLIE 275
QY 394 TLPRIYIIAAVLEFWGGVLLFIQLVPVDYFYFLSIGVLMGKFGITSFMSMLYVFTAEYLP 453
DB 276 KAGRKWILVVLIGTAGSAYFFGTADSLSLLLTAG-VLLSFFNL-GAWGLVYAVTPEQYP 333
QY 454 TLVRNMAVGVTSTASRVGSIIAPYFYVILGAYNMLPYIVMGSLITVLIGIFTL 505
DB 334 TAIATGSGTTAAFGRIIGGIFGLVLTAAARHI-----SFSVIFSFICI 378
```

```
RESULT 9
YAAU_ECOLI
ID YAAU_ECOLI STANDARD; PRT; 443 AA.
AC P31679; P31578; P75628;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN CARB-KEFC INTERGENIC
DE REGION (ORF65/66).
GN YAAU.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
```


Matches 117; Conservative 87; Mismatches 170; Indels 110; Gaps 21.

QY 117 WEFSQDYLSTVTWENLVCDENKWKVPLTTSLEFVGVLGGSFYSGQLSDRFRGR-----170
 :
Db 49 YREEPISPAITLTLWS-----LSVAIFSVMGIGSFGSLGVFNVRGRRNSMLMS 98

QY 171 NVLPATMAVQTSSFLQIFSIHSEM-----FTVLFIIVMGGOISNVYVAFIL 217
 :
Db 99 NILAFLAAVLNMFSGSKM--ALSFMELILGRFIIGLYSGLTGTFGVPMYVGVSPALRGAL 155

QY 218 GTIELGKSVRLIESTLCVCCTFFAVGYMLLPLFAVFI---RD--WRMLLLALTVPGLVCP 272
 :
Db 156 GT-----PHQLGI---VLGILIAQVGLDLINGNDLSWPILLGGIFVFPALLQCI 201

QY 273 LWMFIPESPRW-LISORRFREAEDIIOKAKAMNNATAVPVIFDSVEELNP-LKOQRKAFIL 330
 :
Db 202 ILFPAPESPREFLINRNEENKASVKKL--RGTTDVSSDLQEMKEESQRMREKKVTIM 259

QY 331 DLPETRN-----IATINTINSLLMWLTSYGYPALSALDAPNLHCDAVNLGFLSALIEIPA 384
 :
Db 260 ELFRSPMYROPILAIIVLQSQQSGINAVFYSTS-----FEKSGVGEQP 306

QY 385 YIT-----AWLLRLTPRR-YIIAAVLFWGGVLLFIQLVPVDYY----FL 425
 :
Db 307 YATIGSGVVNTAFTWVSLEFVERAGRRTLHLIAGMACCAILMTIALTLDDQMPWMSYL 366

QY 426 SIGULVMKGFIITSFSAF-----MLYVFTEALYPTLVNRNAVGTSTASRVGS-IIAPIFV 479
 :
Db 367 SIVAII-----EGFAFFEIGPGPIPFWEIABEFLSQGPRAAFVAGVSNWT'SNFIVGMGFQ 422

QY 480 YLGAYNRMLPYIVNGSLTVLIGIFTLF-PPESLGMTLPETLEQMCKVKWPSRKTKRDSM 538
 :
Db 423 YIAQCGSYVEILFTVLLVFFITYFKVETPKGRFDEIAYR-----PROGASQSDK 476

QY 539 ETEE 542
 :
Db 477 TPDE 480

RESULT 13
HXT5_YEAST STANDARD; PRT; 592 AA.
ID HXT5_YEAST AC P38695;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE PROBABLE GLUCOSE TRANSPORTER HXT5.
GN HXT5 OR YHR096C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MC996;
RA Reifemberger E., Koetter P., Ciriacy M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jler M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT *complete nucleotide sequence of Saccharomyces cerevisiae chromosome VIII.*;
RL Science 265:2077-2082(1994).
CC -!- FUNCTION: PROBABLE GLUCOSE TRANSPORTER.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

| RESULT | ID | ENTRY | SEQUENCE | START | END | PROT | LENGTH |
|--------|------------|--|----------|-------|-----|----------------------------|--------|
| 12 | GTRL_CHICK | STANDARD | | 490 | AA | | |
| | ID | GTRL_CHICK | STANDARD | 490 | AA | | |
| | AC | P46896 | | | | | |
| | DT | 01-NOV-1995 (Rel. 32, Created) | | | | | |
| | DT | 01-NOV-1995 (Rel. 32, Last sequence update) | | | | | |
| | DT | 01-NOV-1997 (Rel. 35, Last annotation update) | | | | | |
| | DE | GLUCOSE TRANSPORTER TYPE 1 (GTL1) | | | | | |
| | SL | SLC2A1 OR GLUT1 | | | | | |
| | OS | Gallus gallus (Chicken) | | | | | |
| | OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | | |
| | OC | Archosauria; Aves; Neognathae; Galliformes; Phasianinae; | | | | | |
| | OC | Gallus | | | | | |
| | OX | NCBI_taxID=9031; | | | | | |
| | RN | [1] | | | | | |
| | RN | SEQUENCE FROM N.A. | | | | | |
| | RP | MEDLINE=96157892; PubMed=8589457; | | | | | |
| | RX | Wagstaff P., Kang H.Y., Mylott D., Robbins P.J., White M.K.; | | | | | |
| | RA | "Characterization of the avian GLUT1 glucose transporter: | | | | | |
| | RT | differential regulation of GLUT1 and GLUT3 in chicken embryo | | | | | |
| | RT | fibroblasts." | | | | | |
| | RL | Mol. Biol. Cell 6:1575-1589(1995). | | | | | |
| | CC | -!- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. THIS ISOFORM MAY BE | | | | | |
| | CC | RESPONSIBLE FOR CONSTITUTIVE OR BASAL GLUCOSE UPTAKE. HAS A VERY | | | | | |
| | CC | BROAD SUBSTRATE SPECIFICITY: CAN TRANSPORT A WIDE RANGE OF ALDOSES | | | | | |
| | CC | INCLUDING BOTH PENTOSE AND HEXOSE (BY SIMILARITY). | | | | | |
| | CC | -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. | | | | | |
| | CC | -!- SIMILARITY: HIGH TO OTHER GLUCOSE TRANSPORTERS. | | | | | |
| | CC | -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. | | | | | |
| | CC | ----- | | | | | |
| | CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | | |
| | CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | | |
| | CC | the European Bioinformatics Institute. There are no restrictions on its | | | | | |
| | CC | use by non-profit institutions as long as its content is in no way | | | | | |
| | CC | modified and this statement is not removed. Usage by and for commercial | | | | | |
| | CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/) | | | | | |
| | CC | or send an email to license@isb-sib.ch . | | | | | |
| | CC | ----- | | | | | |
| | EMBL | L07300; AAB02037.1; - | | | | | |
| | DR | InterPro; IPR000803; - | | | | | |
| | DR | InterPro; IPR001066; - | | | | | |
| | DR | InterPro; IPR002439; - | | | | | |
| | DR | Pfam; PF00083; sugar_tr; 1. | | | | | |
| | DR | PRINTS; PR001171; SUGTRNSPORT. | | | | | |
| | DR | PRINTS; PR001172; GLUCTRNSPORT. | | | | | |
| | DR | PRINTS; PR01190; GLUCTRESP01. | | | | | |
| | DR | PROSITE; PS00216; SUGAR_TRANSPORT_1; 1. | | | | | |
| | DR | PROSITE; PS00217; SUGAR_TRANSPORT_2; 1. | | | | | |
| | DR | Duplication; Transmembrane; Sugar transporter; Transport; Glycoprotein; | | | | | |
| | KW | Multigene family. | | | | | |
| | FT | TRANSMEM 12 32 | | | | POTENTIAL. | |
| | FT | DOMAIN 33 65 | | | | EXTRACELLULAR (POTENTIAL). | |
| | FT | TRANSMEM 66 86 | | | | POTENTIAL. | |
| | FT | TRANSMEM 95 115 | | | | POTENTIAL. | |
| | FT | TRANSMEM 126 146 | | | | POTENTIAL. | |
| | FT | TRANSMEM 155 175 | | | | POTENTIAL. | |
| | FT | TRANSMEM 185 205 | | | | POTENTIAL. | |
| | FT | DOMAIN 206 270 | | | | CYTOPLASMIC (POTENTIAL). | |
| | FT | TRANSMEM 271 291 | | | | POTENTIAL. | |
| | FT | TRANSMEM 306 326 | | | | POTENTIAL. | |
| | FT | TRANSMEM 337 357 | | | | POTENTIAL. | |
| | FT | TRANSMEM 368 388 | | | | POTENTIAL. | |
| | FT | TRANSMEM 401 421 | | | | POTENTIAL. | |
| | FT | TRANSMEM 429 449 | | | | | |


```
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE NEG.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; FALSE NEG.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 154 174 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT TRANSMEM 421 441 POTENTIAL.
SQ SEQUENCE 459 AA; 49602 MW; BC8AB53BCB8DD77 CRC64;

Query Match      8.6%; Score 245; DB 1; Length 459;
Best Local Similarity 23.4%; Pred. No. 1.6e-09;
Matches 97; Conservative 81; Mismatches 177; Indels 60; Gaps 17;

QY 134 LYCEDN-----WKVPLTTSLF-----FVGVLGSRFVSGQLSDRFGKKNVLEA 175
DB 34 LVCWSNAVGGULIAQLKALGWTDNSTTATSAITTAGMFLGALVGGIIGDKTGRNAFIL 93
QY 176 TNAVOTGFSFQIIRISWEMETVLFVIVGAGQISNVYVAFILGTEIL-GK-----SVRII 229
DB 94 YEAIHIAKSWVGAAPSPNDFIACRFVGVGGLGALLTLFAGTEYMPGNRGTWSSRVS 153
QY 230 FS--TLGVCFFAVGVMLLPFLAFYFIRDMRLLLALTLPVGLCVPL-WWFIPESRWLI 285
DB 154 FIGNNSYPLCSLIANG--LTPLIS-AEWNMRVQLLIPAILSLIATIALAWRYFPESRWLE 210
QY 286 SORPREAEDIQK-----AARNNTAPAVIFDSVEELNPLKQKAFILDLFRNI--- 338
DB 211 SGRGOEAKVNRSEIEGVIRQTKPLPPVVIADGKAPQAPVYSAITLTVGLKRVILGS 270
QY 339 AITMINSLLMLTVSGVFALSALDAPNLHGDAYLNCFLSALIEIP-AYITAWLLRLTLP 397
DB 271 CVLIAMNVQVTLINWLTPIETQGNLKSIVLNTW-SMGAPGFIETAMLVMBKIPR 328
QY 398 RVIIAAVFWGGVGLVLFQLPVDYFELS--IGLVMLKFCIYSAFMLY-----VFYA 449
DB 329 KFM-----GVGLILIAVGYIYSLQTSMLLTILIGFFLIT--FVYMYCVYASAVYP 379
QY 450 ELYPTLVNRNAVGVSTASRVGSIAPY--FVYLGAYNRMLPYIVMGSLTFLIGI 502
DB 380 EIWPTEAKLRGSLANAVGRISGIAAPYAVAVLSSYGVTVGTFILLGAVSIIVAI 434

RESULT 15
ID KHT2_KLUJA STANDARD: PRT; 566 AA.
AC P53387;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HEXOSE TRANSPORTER 2.
GN KHT2.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAG;
RX MEDLINE=98028406; PubMed=9363776;
RA Weirich J., Goffrin P., Kuger P., Ferrero I., Breunig K.D.;
RT "Influence of mutations in hexose-transporter genes on glucose
RT repression in Kluyveromyces lactis.";
RL Eur. J. Biochem. 243:248-257(1997).
```

Search completed: August 16, 2001, 14:04:00
Job time: 302 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:58:25 ; Search time 36.39 Seconds

(without alignments)
927.935 Million cell updates/sec

Title: US-09-521-195-3

Perfect score: 2883

Sequence: 1 MRDYDEVTAFLGEMGFQRL.....HTRMLKDGQRPRTILKSTAF 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A.Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|----|--------------|
| 1 | 2883 | 100.0 | 557 | 20 | AA1980.DAT.* |
| 2 | 2883 | 100.0 | 557 | 21 | AA1981.DAT.* |
| 3 | 2513 | 87.2 | 557 | 20 | AA1982.DAT.* |
| 4 | 2513 | 87.2 | 557 | 21 | AA1983.DAT.* |
| 5 | 2513 | 87.2 | 557 | 21 | AA1984.DAT.* |
| 6 | 2295.5 | 79.6 | 564 | 21 | AA1985.DAT.* |
| 7 | 2218 | 76.9 | 551 | 20 | AA1986.DAT.* |
| 8 | 2148 | 74.5 | 553 | 20 | AA1987.DAT.* |
| 9 | 2148 | 74.5 | 553 | 21 | AA1988.DAT.* |
| 10 | 785.5 | 27.2 | 560 | 21 | AA1989.DAT.* |
| 11 | 735 | 25.5 | 554 | 19 | AA1990.DAT.* |

| | | | | | |
|----|-------|------|-----|----|--------------|
| 12 | 724.5 | 25.1 | 540 | 22 | AA1991.DAT.* |
| 13 | 719 | 24.9 | 535 | 21 | AA1992.DAT.* |
| 14 | 712 | 24.7 | 556 | 17 | AA1993.DAT.* |
| 15 | 693 | 24.0 | 542 | 21 | AA1994.DAT.* |
| 16 | 690 | 23.9 | 548 | 21 | AA1995.DAT.* |
| 17 | 686.5 | 23.8 | 551 | 21 | AA1996.DAT.* |
| 18 | 673 | 23.3 | 561 | 18 | AA1997.DAT.* |
| 19 | 660.5 | 22.9 | 607 | 21 | AA1998.DAT.* |
| 20 | 656.5 | 22.8 | 536 | 21 | AA1999.DAT.* |
| 21 | 652.5 | 22.6 | 537 | 18 | AA2000.DAT.* |
| 22 | 644 | 22.3 | 545 | 22 | AA2001.DAT.* |
| 23 | 641 | 22.2 | 550 | 21 | AA2002.DAT.* |
| 24 | 638.5 | 22.1 | 551 | 20 | AA2003.DAT.* |
| 25 | 628 | 21.8 | 563 | 20 | AA2004.DAT.* |
| 26 | 607 | 21.1 | 550 | 22 | AA2005.DAT.* |
| 27 | 433 | 15.0 | 480 | 21 | AA2006.DAT.* |
| 28 | 433 | 15.0 | 483 | 21 | AA2007.DAT.* |
| 29 | 380.5 | 13.2 | 515 | 21 | AA2008.DAT.* |
| 30 | 368.5 | 12.8 | 521 | 21 | AA2009.DAT.* |
| 31 | 357.5 | 12.4 | 548 | 21 | AA2010.DAT.* |
| 32 | 344.5 | 11.9 | 397 | 21 | AA2011.DAT.* |
| 33 | 332 | 11.5 | 439 | 22 | AA2012.DAT.* |
| 34 | 327.5 | 11.4 | 378 | 21 | AA2013.DAT.* |
| 35 | 315.5 | 10.9 | 359 | 21 | AA2014.DAT.* |
| 36 | 306.5 | 10.6 | 297 | 21 | AA2015.DAT.* |
| 37 | 302.5 | 10.5 | 520 | 20 | AA2016.DAT.* |
| 38 | 302.5 | 10.5 | 520 | 21 | AA2017.DAT.* |
| 39 | 302.5 | 10.5 | 520 | 21 | AA2018.DAT.* |
| 40 | 302 | 10.5 | 360 | 21 | AA2019.DAT.* |
| 41 | 284 | 9.9 | 339 | 21 | AA2020.DAT.* |
| 42 | 280.5 | 9.7 | 231 | 17 | AA2021.DAT.* |
| 43 | 258 | 8.9 | 742 | 21 | AA2022.DAT.* |
| 44 | 251 | 8.7 | 530 | 21 | AA2023.DAT.* |
| 45 | 234 | 8.1 | 494 | 20 | AA2024.DAT.* |

ALIGNMENTS

RESULT 1
AA1980.DAT.*
ID AA1980.DAT.* standard; Protein: 557 AA.
XX
AC AA1980.DAT.*
XX
DT 23-JUN-1999 (first entry)
XX
DE A protein with cation transporting activity.
XX
KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KW heart disease; cancer; anti-tumour drug; anticancer drug.
XX
OS Homo sapiens.
XX
PN WO9913072-A1.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP04009.
XX
PR 20-MAY-1998; 98JP-0156660.
PR 08-SEP-1997; 97JP-0260972.
XX
(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA Nezu J, Oku A;
XX
DR WPI; 1999-215062/18.
DR N-PSDB; AA206880.
XX
PT Genes homologous with organic cation transporters OCT1 and OCT2,
PT useful in design of new drugs for treatment of diseases due to
PT abnormality of the transporter functions

Murine organic anion
Rat liver anion tr
Rat OCT-1 protein.
Human cerebral org
A human organic an
A human organic an
Human osteoclast t
Hydrophobic domain
Rat cerebral organ
Mouse osteoclast t
Mouse organic anio
Human organic anio
Rat organic anion
Human organic anio
Human organic anio
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human organic cati
Arabidopsis thalia
Corynebacterium gl
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human organic cati
Human saccharide-t
Human secreted pro
Cat flea HMT synap
Glucose transporte

```

XX PS Claim 1: Page 51-55; 97pp; Japanese.
XX CC The present sequence represents a protein with cation transporting
XX CC activity. The genes are significantly homologous with organic cation
XX CC transporters OCT1 and OCT2. The genes may be used in drug development,
XX CC particularly in the treatment of diseases due to abnormality of the
XX CC organic cation transporter functions e.g. fatty liver, heart diseases
XX CC and cancers, by controlling such as by inhibition or activation.
XX CC Administration of anti-tumour and anticancer drugs in combination with
XX CC a transporter protein inhibiting agent allows the agents to penetrate
XX CC into the diseased cells to enhance the drug action.
XX SQ Sequence 557 AA;

Query Match 100.0%; Score 2883; DB 20; Length 557;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEMGPFORLIFLLSASIPNGFTGLSSVFLIATPEHRCRVDPDAANLSS 60
DB 1 mrdydevtaflgwgpfqrlifllsasilpngftglssvfliatpehrcrvpdaanlss 60

QY 61 AWRNHTVPLRLDREVPHSRCRRYRLATIANFSALGLEPGRDVLQLESCLDGWEFS 120
DB 61 awrnhtvplrlrdgrevphsrryrlatiansalglepgrdvlqleescldgwefs 120

QY 121 QDVYLSTIVTEWNLVCEDDWKAPLTISLFFVGLGSGFISGQLSDRFGRKNVLFVTMGQM 180
DB 121 qdvylstivtewnlvceddwkapltislfvfgllsgsfisgqlsdrfgrknvlfvtmgqm 180

QY 181 TGFSELOIFSKNFEMFVFLVVGMOISNYVAAFVLGTEILGKSVRIIPSTLGVCIIFYA 240
DB 181 tgfsefioifsknfemfvflvvgmoisnyvaafvlgteilgksvriipstlsgvcifya 240

QY 241 FGMYVPLFAYFIRDRMLLVALTMPGVLVCMWFIPESPRWLISQGRFEAEVIRKA 300
DB 241 fgyvplfayfirdrmlvaltmpgvlvcmwfiipesprwlisqgrfeaevirka 300

QY 301 AKANGIVVPSTIFDSELDLSSKKQSHNLDLRLTWNIRMTIMSLMWTISVGYFG 360
DB 301 akangivvpstifdseldlsskkqshnldlrltwnirmtimslmwtisvgyfg 360

QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLQLQYLPRLRYSMATALLFLGSGVLLFMQLV 420
DB 361 lsltdpnlhgdifvncflsamvevpayvlawlqlqylprlrysmatallflgsgvllfmqlv 420

QY 421 PPDLYYLATVLMVGKFGVTAAFSMVYVYTAELYPTVVRNMGVSVSTASRLGSLSPYF 480
DB 421 ppdllylatvlmvgkfgvtaafsmvyvyytaelyptvvrnmgvsvstasrlgslsypf 480

QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDIDQMLRVKGMKHKRTPSHTR 540
DB 481 vylgaydrflpyilmgsltilitaitlflpesfgtlpdtdqmlrvkgmkhkrtpshtr 540

QY 541 MLKDGQERTILKSTAF 557
DB 541 mlkdqgertilkstaf 557

RESULT 2
ID AAY83929 standard; Protein; 557 AA.
XX AC AAY83929;
XX DT 05-JUL-2000 (first entry)
XX DE Human carnitine transporter protein OCTN2.
KW Organic cation transportation; human; carnitine transporter protein;
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;

```

```

KW juvenile visceral steatosis.
XX OS Homo sapiens.
XX PN WO200014210-A1.
XX PD 16-MAR-2000.
XX PF 07-SEP-1999; 99WO-JP04853.
XX PR 07-SEP-1998; 98JP-0252683.
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX PI Nezu J, Oku A;
XX WI: 2000-256966/22.
XX N-PSDB; AAA03889.
XX PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
XX PT cation transporter, useful as diagnostic tool
XX PS Example 1; Page 36-40; 106pp; Japanese.
XX CC This sequence represents the human carnitine transporter protein OCTN2.
XX CC The coding sequence can be used as a target for diagnosis of systemic
XX CC carnitine deficiency by detecting the presence of mutations in the
XX CC sequence, especially seen in the disease juvenile visceral steatosis
XX CC (jvs). The wild type OCTN2 gene can be used in the gene therapy of the
XX CC disease state.
XX SQ Sequence 557 AA;

Query Match 100.0%; Score 2883; DB 21; Length 557;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEMGPFORLIFLLSASIPNGFTGLSSVFLIATPEHRCRVDPDAANLSS 60
DB 1 mrdydevtaflgwgpfqrlifllsasilpngftglssvfliatpehrcrvpdaanlss 60

QY 61 AWRNHTVPLRLDREVPHSRCRRYRLATIANFSALGLEPGRDVLQLESCLDGWEFS 120
DB 61 awrnhtvplrlrdgrevphsrryrlatiansalglepgrdvlqleescldgwefs 120

QY 121 QDVYLSTIVTEWNLVCEDDWKAPLTISLFFVGLGSGFISGQLSDRFGRKNVLFVTMGQM 180
DB 121 qdvylstivtewnlvceddwkapltislfvfgllsgsfisgqlsdrfgrknvlfvtmgqm 180

QY 181 TGFSELOIFSKNFEMFVFLVVGMOISNYVAAFVLGTEILGKSVRIIPSTLGVCIIFYA 240
DB 181 tgfsefioifsknfemfvflvvgmoisnyvaafvlgteilgksvriipstlsgvcifya 240

QY 241 FGMYVPLFAYFIRDRMLLVALTMPGVLVCMWFIPESPRWLISQGRFEAEVIRKA 300
DB 241 fgyvplfayfirdrmlvaltmpgvlvcmwfiipesprwlisqgrfeaevirka 300

QY 301 AKANGIVVPSTIFDSELDLSSKKQSHNLDLRLTWNIRMTIMSLMWTISVGYFG 360
DB 301 akangivvpstifdseldlsskkqshnldlrltwnirmtimslmwtisvgyfg 360

QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLQLQYLPRLRYSMATALLFLGSGVLLFMQLV 420
DB 361 lsltdpnlhgdifvncflsamvevpayvlawlqlqylprlrysmatallflgsgvllfmqlv 420

QY 421 PPDLYYLATVLMVGKFGVTAAFSMVYVYTAELYPTVVRNMGVSVSTASRLGSLSPYF 480
DB 421 ppdllylatvlmvgkfgvtaafsmvyvyytaelyptvvrnmgvsvstasrlgslsypf 480

QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDIDQMLRVKGMKHKRTPSHTR 540
DB 481 vylgaydrflpyilmgsltilitaitlflpesfgtlpdtdqmlrvkgmkhkrtpshtr 540

```

QY 541 MLKDGQERTILKSTAF 557
DB 541 mlkdqgertilkstaf 557
RESULT 3
ID AAY01652 standard; Protein; 557 AA.
AC AAY01652;
DT 23-JUN-1999 (first entry)
XX A protein with cation transporting activity.
DE A protein with cation transporting activity.
KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KW heart disease; cancer; anti-tumour drug; anticancer drug.
XX Mus musculus.
XX WO9913072-A1.
XX 18-MAR-1999.
XX PD
XX PF 07-SEP-1998; 98WO-JP04009.
XX PR 20-MAY-1998; 98JP-0156660.
XX PR 08-SEP-1997; 97JP-0260972.
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Nezu J, Oku A;
XX WPI; 1999-215062/18.
XX N-PSDB; AAX26902.
XX Genes homologous with organic cation transporters OCT1 and OCT2,
PT useful in design of new drugs for treatment of diseases due to
PT abnormality of the transporter functions
XX
XX Claim 1; Page 75-79; 97pp; Japanese.
XX The present sequence represents a protein with cation transporting
CC activity. The genes are significantly homologous with organic cation
CC transporters OCT1 and OCT2. The genes may be used in drug development,
CC particularly in the treatment of diseases due to abnormality of the
CC organic cation transporter functions e.g. fatty liver, heart diseases
CC and cancers, by controlling such as by inhibition or activation.
CC Administration of anti-tumour and anticancer drugs in combination with
CC a transporter protein inhibiting agent allows the agents to penetrate
CC into the diseased cells to enhance the drug action.
XX
SQ Sequence 557 AA;

Query Match 87.2%; Score 2513; DB 20; Length 557;
Best Local Similarity 85.5%; Pred. No. 1.9e-265;
Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;
QY 1 MRDYDEVTAFLGEMGPORLIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
DB 1 mrdydevtaflgwgptqrlifflsasipngftglssvflatpchrclvphtvnlss 60
QY 61 AWRNHTVPLRLDRGVPVPHSCRRYRIATIANFSALEPGRDVDLQQLQESCLDQWEPF 120
DB 61 awrnhtvplrlkgrvqpqkrryrlatianselglegrdvrdieqlqesclgweyjd 120
QY 121 ODVLSPTVTEWNLVCEDDWKAPLTLISLFFVGVLLGSFTISGQLSDRFRKKNVLFVTMGMO 180
DB 121 kdvlstvtewdlvckddwkapltslffvgvlmgfsgqlsdrfgrknvlfmgmq 180
QY 181 TGFSFLQIFSKFEMFVFLVGVGMQISNYAAVFLGTTEILGKSVRIIFSLIGVCIFYA 240

DB 181 tgfsflqvsvnfemftvlflvgmqgisnyvaafvlgteilsksrliifatlvcifya 240
QY 241 FGYMWLPFLFAYFTRDWRMLLVALTPGVLVCVALWFIPESPRWLIISQGRFEAEVIRKA 300
DB 241 fgmvlplfayfirdwrmlllaltvpqvlcgalwffipesprwlisqgrikeaeavirka 300
QY 301 AKANGIVVPSTIFDPSLODLSKQOOSHNLIDLRTNIRMTIMSLWMTISVGYFG 360
DB 301 akingivvapstifdpselqdnstkpqlhnydlirnrivritmsilwltisvgyfg 360
QY 361 LSLETPNLHGDIFVNCFLSAMVEVPAYLAWLLLOVLPFRYSMATALLFGGSVLLPQVLV 420
DB 361 lsldtpnlhgdlyvncflaavevpaylawlllqylpryysaallfggsvllmqvlv 420
QY 421 PPDIYYIATVLMVWGKFCVTAAFSMVYVYVYAEIYPTVVRNMGVGVSTASRLSILSPYF 480
DB 421 pselfyistalvmvgkfgitsaysmvvytaelyptvvrnmvgvstasrlsilsyfi 480
QY 481 VYLGAYDRFLPYILMGSILITLITLTLFLPESFGTLPDIDQMLRVKMKHRTSPSHTR 540
DB 481 vyigaydrfpyilmgslitlittaltiffesfgvplpdtidqmlrvkqkqwgqsgtr 540
QY 541 MLKDGQERTILKSTAF 557
DB 541 mqkdgeesptvlkstaf 557
RESULT 4
ID AAB20580 standard; Protein; 557 AA.
XX AAB20580;
XX 11-DEC-2000 (first entry)
XX Mouse OCTN2 amino acid sequence.
XX Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
KW identification; regulator; carnitine transport.
XX Mus musculus.
XX WO200046368-A1.
XX 10-AUG-2000.
XX 04-FEB-2000; 2000WO-JP00619.
XX 05-FEB-1999; 99JP-0028406.
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Nezu J, Ose A;
XX WPI; 2000-586982/55.
XX Organic cation transporter gene OCTN3 expressed in testis for
PT identification of regulators of carnitine transport for use as drugs
XX Example 2; Fig 2; 58pp; Japanese.
XX The present invention describes a mouse organic cation transporter
CC protein (OCTN3). Also described are: (1) a method for screening
CC compounds for their ability to regulate the transport of an organic
CC cation into the cell, by generating a cell expressing OCTN3 at the
CC cell membrane, contacting with the compound and organic cation, and
CC observing the degree of transport of the organic cation; and (2) a
CC method for screening compounds for their ability to be transported into
CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
CC membrane, contacting with the compound and observing the degree of
CC transport of the compound. OCTN3 can be used for the identification of
CC regulators of the transport of organic cations (especially carnitine)

CC into cells by OCTN3, for use as drugs. The present sequence represents
CC the mouse OCTN2 amino acid sequence, which is used in an example from
CC the present invention.

XX Sequence 557 AA;

Query Match 87.2%; Score 2513; DB 21; Length 557;
Best Local Similarity 85.5%; Pred. No. 1.9e-265;
Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYEVTAFELGEGPQRIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
DB 1 mrdydevtafelgwegpqrliifllsasiipngfngmsivflagtpehrcrlvhtnvlss 60
QY 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALGLEGRDVLGQLESCLDGWEFS 120
DB 61 awrnhsipletkdgrrvpqkrryrlatianselglegrdvldleqesclgdgweyd 120
QY 121 QDVYLSITVTENWLVCDDWKAPLTISLFFVGLGSGFISGQSLDRGRKNVLFVTMGMO 180
DB 121 kdvlstivtwedlvcddwkaptltslffvglmgsfsgqlsdrgrknvlfmgmq 180
QY 181 TGFSEFLQIFSKNFMFVFLVGMGOISNYAAAFVLGTEILGKSVRIIFSTLGVCIIFYA 240
DB 181 tgfsflqifsvnfemftvflvvgmqisnyaaafvlgteilsksrilifatlvgcifya 240
QY 241 FGYMWLPFLAFYFIRDRMMLVALTMPCVLCVALWFWIPESPRWLISQGRFEAEVIIRKA 300
DB 241 fgmwlpflafyfirdrwmllaltvpcvlgcalwfwipesprwlisqgrfkaeeviirka 300
QY 301 AKANGIVVPSTIFDPSELQDLSSKKQSHNILDLLRTNIRMTVIMSIMLWMTISVGYFG 360
DB 301 akingivapstifdpseqlqdlstskpqlhnydlrtnrirvmtimsilwltisvgyfg 360
QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLQLYLPRLYSMATALFLGGSVLLFMQLV 420
DB 361 lslatpnlhgdlyncflsamvevpayvawlllqlprylsaaflggsvllfmqlv 420
QY 421 PPDLXYLATVLMVGRFGVTAAFSMYVYVTAELYPTVVRNMGVSGVSTASRLGSIILSPYF 480
DB 421 pselfylstalvmvgrfgvtaafsmvyyvtaelyptvvrnmvgvsgvstasrlgsilspyf 480
QY 481 VYLGAYDRFLPYILMGSLLTILTAITLFLPESFGTLPDIDQMLRVKGMKHKRTPSHTR 540
DB 481 vylgaydrflpyilmgslltaltliffpesfgvplpdtidqmlrvkgmkhkrtpshtr 540
QY 541 MLKDGQERPTILKSTAF 557
DB 541 mkgdgeesptvilkstaf 557

RESULT 5

AA83930
ID AAY83930 standard; Protein; 557 AA.

AC AAY83930;

XX 05-JUL-2000 (first entry)

DE Mouse carnitine transporter protein OCTN2.

KW Organic cation transportation; mouse; carnitine transporter protein;
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
KW juvenile visceral steatosis.

OS Mus musculus.

PN WO200014210-A1.

XX 16-MAR-2000.

XX 07-SEP-1999; 99WQ-JP04853.

PF

XX 07-SEP-1998; 99JP-0252683.
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Nezu J, Oku A;

XX WPI; 2000-256966/22.
XX N-PSDB; AAA09890.

PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
PT cation transporter, useful as diagnostic tool
PS Example 1; Page 47-51; 106pp; Japanese.

XX This sequence represents the mouse carnitine transporter protein OCTN2.
CC The coding sequence of the corresponding human protein can be used as a
CC target for diagnosis of systemic carnitine deficiency by detecting the
CC presence of mutations in the sequence, especially seen in the disease
CC juvenile visceral steatosis (jvs). The wild type OCTN2 gene can be used
CC in the gene therapy of the disease state.

SQ Sequence 557 AA;

Query Match 87.2%; Score 2513; DB 21; Length 557;
Best Local Similarity 85.5%; Pred. No. 1.9e-265;
Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYEVTAFELGEGPQRIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
DB 1 mrdydevtafelgwegpqrliifllsasiipngfngmsivflagtpehrcrlvhtnvlss 60
QY 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALGLEGRDVLGQLESCLDGWEFS 120
DB 61 awrnhsipletkdgrrvpqkrryrlatianselglegrdvldleqesclgdgweyd 120
QY 121 QDVYLSITVTENWLVCDDWKAPLTISLFFVGLGSGFISGQSLDRGRKNVLFVTMGMO 180
DB 121 kdvlstivtwedlvcddwkaptltslffvglmgsfsgqlsdrgrknvlfmgmq 180
QY 181 TGFSEFLQIFSKNFMFVFLVGMGOISNYAAAFVLGTEILGKSVRIIFSTLGVCIIFYA 240
DB 181 tgfsflqifsvnfemftvflvvgmqisnyaaafvlgteilsksrilifatlvgcifya 240
QY 241 FGYMWLPFLAFYFIRDRMMLVALTMPCVLCVALWFWIPESPRWLISQGRFEAEVIIRKA 300
DB 241 fgmwlpflafyfirdrwmllaltvpcvlgcalwfwipesprwlisqgrfkaeeviirka 300
QY 301 AKANGIVVPSTIFDPSELQDLSSKKQSHNILDLLRTNIRMTVIMSIMLWMTISVGYFG 360
DB 301 akingivapstifdpseqlqdlstskpqlhnydlrtnrirvmtimsilwltisvgyfg 360
QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLQLYLPRLYSMATALFLGGSVLLFMQLV 420
DB 361 lslatpnlhgdlyncflsamvevpayvawlllqlprylsaaflggsvllfmqlv 420
QY 421 PPDLXYLATVLMVGRFGVTAAFSMYVYVTAELYPTVVRNMGVSGVSTASRLGSIILSPYF 480
DB 421 pselfylstalvmvgrfgvtaafsmvyyvtaelyptvvrnmvgvsgvstasrlgsilspyf 480
QY 481 VYLGAYDRFLPYILMGSLLTILTAITLFLPESFGTLPDIDQMLRVKGMKHKRTPSHTR 540
DB 481 vylgaydrflpyilmgslltaltliffpesfgvplpdtidqmlrvkgmkhkrtpshtr 540
QY 541 MLKDGQERPTILKSTAF 557
DB 541 mkgdgeesptvilkstaf 557

RESULT 6

AAB20578
ID AAB20578 standard; Protein; 564 AA.

XX AAB20578;
 XX 11-DEC-2000 (first entry)
 XX Mouse OCTN3 protein SEQ ID NO:1.
 DE Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
 KW identification; regulator; carnitine transport.
 XX Mus musculus.
 OS
 PN W0200046368-A1.
 XX 10-AUG-2000.
 XX 04-FEB-2000; 2000WO-JP00619.
 XX 05-FEB-1999; 99JP-0028406.
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX Nezu J, Ose A;
 XX WPI; 2000-586982/55.
 XX N-PSDB; AAA88053.
 XX Organic cation transporter gene OCTN3 expressed in testis for
 PT identification of regulators of carnitine transport for use as drugs
 PT
 XX Claim 1; Page 34-39; 58pp; Japanese.
 XX The present invention describes a mouse organic cation transporter
 CC protein (OCTN3). Also described are: (1) a method for screening
 CC compounds for their ability to regulate the transport of an organic
 CC cation into the cell, by generating a cell expressing OCTN3 at the
 CC cell membrane, contacting with the compound and organic cation, and
 CC observing the degree of transport of the organic cation; and (2) a
 CC method for screening compounds for their ability to be transported into
 CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
 CC membrane, contacting with the compound and observing the degree of
 CC transport of the compound. OCTN3 can be used for the identification of
 CC regulators of the transport of organic cations (especially carnitine)
 CC into cells by OCTN3, for use as drugs. The present sequence represents
 CC mouse OCTN3.
 XX
 SQ Sequence 564 AA;

Query Match 79.6%; Score 2295.5; DB 21; Length 564;
 Best Local Similarity 80.6%; Pred. No. 1.1e-241;
 Matches 435; Conservative 54; Mismatches 48; Indels 3; Gaps 1;
 1 MRDYDEVTAFLGEGWGPFORLIFFLLSASIPNGFTGLSVFLIATPEHRCRVPDAANLSS 60
 1 mldydevtaflgwggtfqrllflllsasipngftglsvflaipaehrcrripdvtvlss 60
 61 AWRNHTVPLRGDREVPHSRRYRLATIANFSALGLEPGRDVLGLQLEQSCLDGWEFS 120
 61 awrnhsipmetkdqpevpqckryrlatiantfsejglepgrdvdleqleqncldgweyd 120
 121 ODVYLSITVTENLVCEDDWKAPLTSFFVGVLLGSFTSGQLSDRFGRKNVLFVTMGQ 180
 121 kdilsvitvewdlvcddwkdkapltstffvvgvllgsfsgqlsdrgfgrknlfvltmnh 180
 181 TGFSEFLQIFSKNFEMFVFLVGVGMQISNYAAAFVLGTEILGKSVRIIFSTLGVCFYA 240
 181 tgfseflqvsvnfemflltytlvgmhisnyaaafvlgtemlksvriifatlvgvciffa 240
 241 FGWVPLPLFAYFTRDWRMLLVALTWPGVLCVALMWFIPESPRWLISQGRFEAEVIIRKA 300
 241 fgfmvplplfayfirewrrillaitlpgvlgcalwvfipesprwlisqgrikeaeavirka 300

QY 301 AKANGIVVPSTIFDPSE---LQDLSSKKQOSHNLDLLRTNIRMTVIMSLMWTISVG 357
 DB 301 akingivapstifopsetnklqddsskpkshhlydvrtpniriltimsiilwtisvg 360
 QY 358 YFGLSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLLOQLPFRYSMATALFLGGSVLLFM 417
 DB 361 yfglsldtpnlnngnyvncflaavevpayvlawlllqhvsrrysmagsiflgssvllv 420
 QY 418 QLVPPDIYYLATVLMVGKFGVTAAFSMVYVYTAELIPTVVRNMGVGVSTASRLGSILS 477
 DB 421 qlvpsdihylsttlvmvgkfgitsaysmvvyvtaelyptvvrnmgvgsstasrlgsils 480
 QY 478 PYFVYLGAIDRELPYILMGSILTILTAITLFLPESFGTLPDITDQMLRVKGMKHKRTPS 537
 DB 481 pyfvyldaydrilpyllmgsiltitaitiffessgvsipetidmqkvkikqrqls 540
 RESULT 7
 AAY01649
 ID AAY01649 standard; Protein; 551 AA.
 XX
 AC AAY01649;
 XX
 DT 23-JUN-1999 (first entry)
 XX
 DE A protein with cation transporting activity.
 XX
 KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
 KW heart disease; cancer; anti-tumour drug; anticancer drug.
 XX
 OS Homo sapiens.
 XX
 FN W09913072-A1.
 XX
 PD 18-MAR-1999.
 XX
 PF 07-SEP-1998; 98WO-JP04009.
 XX
 PR 20-MAY-1998; 98JP-0156660.
 PR 08-SEP-1997; 97JP-0260972.
 XX
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX
 PI Nezu J, Oku A;
 XX
 DR WPI; 1999-215062/18.
 DR N-PSDB; AAX26879.
 XX
 PT Genes homologous with organic cation transporters OCT1 and OCT2,
 PT useful in design of new drugs for treatment of diseases due to
 PT abnormality of the transporter functions
 XX
 PS Claim 1; Page 41-45; 97pp; Japanese.
 XX
 CC The present sequence represents a protein with cation transporting
 CC activity. The genes are significantly homologous with organic cation
 CC transporters OCT1 and OCT2. The genes may be used in drug development,
 CC particularly in the treatment of diseases due to abnormality of the
 CC organic cation transporter functions e.g. fatty liver, heart diseases
 CC and cancers, by controlling such as by inhibition or activation.
 CC Administration of anti-tumour and anticancer drugs in combination with
 CC a transporter protein inhibiting agent allows the agents to penetrate
 CC into the diseased cells to enhance the drug action.
 XX
 SQ Sequence 551 AA;

Query Match 76.9%; Score 2218; DB 20; Length 551;
 Best Local Similarity 75.9%; Pred. No. 3.1e-233;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;
 1 MRDYDEVTAFLGEGWGPFORLIFFLLSASIPNGFTGLSVFLIATPEHRCRVPDAANLSS 60
 1 mldydevtaflgwggtfqrllflllsasipngftglsvflaipaehrcrripdvtvlss 60

Db 1 mrdydeviaflegwqfgrlifflllsasllpngfngmsvflagtphehrcrvpdaanlss 60
 QY 61 AWRNHTVPLRLRDGREGVPHSCRRYRLATIANFSALGLEPGRDVLGQLEQESCLDGMWFS 120
 Db 61 avrnsvplrlrdgrevphscsryrlatiansalglepgrdvdldglegescldgwefs 120
 QY 121 QDYLSTIVTENLVCEDDWKAPLTISLFFVGVLLGSLISGOLSDRFRGKNVLFVTMGQ 180
 Db 121 qdylstivtenlvceadnwkwpltslffvgvllgsvsgqlsdrfrgknvlfatmavq 180
 QY 181 TGFSLQIFSKNFEMFVFLVGVGMQISNVAAAFVLGTEILGKSVRIIFSTLGVCIPIYA 240
 Db 181 tgfslqifsknfemfvlvfgvmgqisnyvvaflgtellgksvriifstlgtvctffa 240
 QY 241 FGMYLPLFAYFIRDRWMLLVALTMGPVLCVALMWFIPESPRWLISQGRFEEAEVIIRKA 300
 Db 241 vgymlplfayfirdwrmllaltvgvlpwlfipesprwlisqrfrfreaedliqka 300
 QY 301 AKANGIVPSTTFDSELDLSSKKQSHNILDLLRTWIRMVIMSLMWTISVGYFG 360
 Db 301 akmntavpavifd--sveelnplkqgkafildlfrtniaimtimslmltswgyfa 358
 QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLLOYLPRRYSMATALFLGGSVLLFMQLV 420
 Db 359 lslapnldhdaylncflsalieipayitawlrltpryiaavlfwggvllfiqlv 418
 QY 421 PPDLYLATVLMVGVKFGVTAAFSMVYVYTAELYPVVRNMGVSVSTASRLGSLSPYF 480
 Db 419 pvdyyflslglvmlgkfgitsafmsllyftaelypvlvrnmavgtstasrvgsliaapyf 478
 QY 481 VYLGYDRFLPVLGSLTILTAITLFLPESFGTLPDITDQMLRVKGMKHKRTPSHTR 540
 Db 479 vylgaydrflpvlmgsltlvlgiflfpeslgnltplietlegmkvkwfrsgk---ktr 535
 QY 541 MLKDGQERPTILKSTAF 557
 Db 536 dsmeteepkvl-itaaf 551

RESULT 8
 ID AAY01651
 AC AAY01651;
 XX AAY01651;
 DT 23-JUN-1999 (first entry)
 XX A protein with cation transporting activity.
 DE Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
 KW heart disease; cancer; anti-tumour drug; anticancer drug.
 XX Mus musculus.
 OS WO9913072-A1.
 PN 18-MAR-1999.
 XX 07-SEP-1998; 98WQ-JP04009.
 PF 20-MAY-1998; 98JP-0156660.
 PR 08-SEP-1997; 97JP-0260972.
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA Nezu J, Oku A;
 PI WPI: 1999-215062/18.
 DR N-FSDB; AAX26898.
 XX Genes homologous with organic cation transporters OCT1 and OCT2,
 PT useful in design of new drugs for treatment of diseases due to
 PT abnormality of the transporter functions

XX Claim 1; Page 63-67; 97pp; Japanese.
 PS The present sequence represents a protein with cation transporting
 CC activity. The genes are significantly homologous with organic cation
 CC transporters OCT1 and OCT2. The genes may be used in drug development,
 CC particularly in the treatment of diseases due to abnormality of the
 CC organic cation transporter functions e.g. fatty liver, heart diseases
 CC and cancers, by controlling such as by inhibition or activation.
 CC Administration of anti-tumour and anticancer drugs in combination with
 CC a transporter protein inhibiting agent allows the agents to penetrate
 CC into the diseased cells to enhance the drug action.
 XX SQ Sequence 553 AA;

Query Match 74.5%; Score 2148; DB 20; Length 553;
 Best Local Similarity 72.2%; Pred. No. 1.4e-225;
 Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;

QY 1 MRDYDEVIAFLEGWQFGRLIFFLLSASLLPNGFNGMSVFLAGTPHEHRCRVDPDAANLSS 60
 Db 1 mrdydeviaflegwqfgrliffllsasiipngfngmsvflagtphehrcrvdpdtnlss 60
 QY 61 AWRNHTVPLRLRDGREGVPHSCRRYRLATIANFSALGLEPGRDVLGQLEQESCLDGMWFS 120
 Db 61 swrnhsipltkdgrgvppqscrryrlatiansamlepggdvldleqesclgdgweyd 120
 QY 121 QDYLSTIVTENLVCEDDWKAPLTISLFFVGVLLGSLISGOLSDRFRGKNVLFVTMGQ 180
 Db 121 kdlfstivtenlvceadnwkwpltslffvgvllgsvsgqlsdrfrgknvlfatmavq 180
 QY 181 TGFSLQIFSKNFEMFVFLVGVGMQISNVAAAFVLGTEILGKSVRIIFSTLGVCIPIYA 240
 Db 181 tgfslqifsknfemfvlvfgvmgqisnyvvaflgtellgksvriifstlgtvctffa 240
 QY 241 FGMYLPLFAYFIRDRWMLLVALTMGPVLCVALMWFIPESPRWLISQGRFEEAEVIIRKA 300
 Db 241 igymvlpifayfirdwrmllaltlpglfcvplwlfipesprwlisqrfrfreaedliqka 300
 QY 301 AKANGIVPSTTFDSELDLSSKKQSHNILDLLRTWIRMVIMSLMWTISVGYFG 360
 Db 301 akmslvapagifdplelqelskqkvilldlfrtniatitvmaavlmltswgyfa 360
 QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLLOYLPRRYSMATALFLGGSVLLFMQLV 420
 Db 361 lslnvpnlhgdvylncflslgievpayftawlrltpryiaavlfwggvllfiqvv 420
 QY 421 PPDLYLATVLMVGVKFGVTAAFSMVYVYTAELYPVVRNMGVSVSTASRLGSLSPYF 480
 Db 421 pedyntvsglvmglkfgitsafmsllyftaelypvlvrnmavgtstasrvgsliaapyf 480
 QY 481 VYLGYDRFLPVLGSLTILTAITLFLPESFGTLPDITDQMLRVKGMKHKRTPSHTR 540
 Db 481 vylgaydrflpvlmgsltlvlgiflfpesfgtvlpenlegmkvkrfgcgk---kst 537
 QY 541 MLKDGQERPTILKSTAF 557
 Db 538 vsvdreespkvl-itaaf 553

RESULT 9
 AAB20579
 ID AAB20579 standard; Protein; 553 AA.
 XX AAB20579;
 AC AAB20579;
 DT 11-DEC-2000 (first entry)
 XX Mouse OCTN1 amino acid sequence.
 DE Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
 KW identification; regulator; carnitine transport.

OS Mus musculus.
 PN WO200046368-A1.
 XX 10-AUG-2000.
 XX 04-FEB-2000; 2000WO-JP00619.
 XX 05-FEB-1999; 99JP-0028406.
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX Nezu J, Ose A;
 XX WPI; 2000-586982/55.
 XX Organic cation transporter gene OCTN3 expressed in testis for
 PT identification of regulators of carnitine transport for use as drugs
 XX Example 2; Fig 2; 58pp; Japanese.
 XX The present invention describes a mouse organic cation transporter
 CC protein (OCTN3). Also described are: (1) a method for screening
 CC compounds for their ability to regulate the transport of an organic
 CC cation into the cell, by generating a cell expressing OCTN3 at the
 CC cell membrane, contacting with the compound and organic cation, and
 CC observing the degree of transport of the organic cation; and (2) a
 CC method for screening compounds for their ability to be transported into
 CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
 CC membrane, contacting with the compound and observing the degree of
 CC transport of the compound. OCTN3 can be used for the identification of
 CC regulators of the transport of organic cations (especially carnitine)
 CC into cells by OCTN3, for use as drugs. The present sequence represents
 CC the mouse OCTN1 amino acid sequence, which is used in an example from
 XX the present invention.
 XX Sequence 553 AA;
 SQ
 Query Match 74.5%; Score 2148; DB 21; Length 553;
 Best Local Similarity 72.2%; Pred. No. 1.4e-225;
 Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;
 QY 1 MRDYDEVTAFLEWGFQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDANLSS 60
 Db 1 mrdydeviafigewgftqrlifllsasipngfmgmsvvlagtpchrclvpdtnlss 60
 QY 61 AWRNHTVPLRLDRGVPVSCRRYRLATIANFSALEPGRVDLGLQESCLDGWFS 120
 Db 61 swrnhsiptekdgrvqpsccrryrlatiansfamlpggqdvdlqeqescldgweyd 120
 QY 121 QDYLSTIVTEWNLVCEDDKAPLTLISLFGVLLGSFTISGQLSDRFGKKNLVFTMGMQ 180
 Db 121 kdiflstivtewnlvceddkwtltslffvgvllgsftisgqlsdrfgrkknlfatmaq 180
 QY 181 TGFSELOIFSKNEFMFVFLVINGMGQISNYAAFLVGLTEILCKSVRIIFSLGVCIFA 240
 Db 181 tgrsfvqlftsnwemtvflvaygmqisnyvavfllgteilsksvriifstlgvctffa 240
 QY 241 FGVMVLPFLFAFYTRDMRLMLVALTMPGLCVLWMPFIPSPRWLISQGRFEEAEVIRKA 300
 Db 241 igymvlpflfayfirdwmrlmlaltpglfcvplwfpesprwlisqrfaeeaiqlka 300
 QY 301 AKANGIVVPTIFDPSELQSLSSKKQOSHNLDLRLTNIRMTIMSIMLWMTISVGYGF 360
 Db 301 akmsivapagifdplelqelskqkvildlfrtniatitvmaavlmltsvgyfa 360
 QY 361 LSLDTPNLHGDIFWNCFLSMEVPAVLANLLQYLPRYSMATALFLGGSVLLPMQLV 420
 Db 361 lslnvpnlhgdvynclflsliuevayftawlilrlpryiaagiylfvgggvllliqv 420
 QY 421 PPDLYLATLVLMVGFGVTAFAFSMVVYVYTAELYPTVWRNMGVGSSTASRLGSLSPYF 480

Db 421 pedynfvsvlgmkgkigtsafsmlyftaelyptivrmavgitsmasrvgsilapyf 480
 QY 481 VYLGAIDRFLPYILMGSITLITAILTLFLPESFTPLPDTIDQMLRVKGMKHKRTPSHTR 540
 Db 481 vylgaynrllpyilmsltvligititffpesfgvtipenleqmkgvrgfcgk---kst 537
 QY 541 MLKDGQERPTILKSTAF 557
 Db 538 vsvreespkvl-ita 553
 RESULT 10
 AAB43038
 ID AAB43038 standard; Protein; 560 AA.
 XX AAB43038;
 XX 08-FEB-2001 (first entry)
 XX Human ORFX ORF2802 polypeptide sequence SEQ ID NO:5604.
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX Homo sapiens.
 OS
 XX WO200058473-A2.
 XX 05-OCT-2000.
 XX 31-MAR-2000; 2000WO-US08621.
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 XX Shimkets RA, Leach M;
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC77247.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PS Claim 11; Page 4791-4792; 5507pp; English.
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating

Db 458 ptfvngvmvcsledigdiptfivfrlrevwqalpdlfavlgllaagvtlllpctk 517
 QY 514 GTPLPDTI 521
 Db 518 gvalpetm 525

RESULT 12
 AAB49401
 ID AAB49401 standard; Protein; 540 AA.
 XX
 AC AAB49401;
 XX
 DT 02-MAR-2001 (first entry)
 XX
 DE Murine organic anion transporter 6.
 XX
 KW Murine; organic anion transporter 6; mOATP6; cancer; inflammation;
 KW cardiovascular disease; central nervous system disorder; kidney disease;
 KW liver disease; autoimmune disease.
 XX
 OS Mus sp.
 XX
 PN WO200070048-A1.
 XX
 PD 23-NOV-2000.
 XX
 PF 15-MAY-2000; 2000WO-US13316.
 XX
 PR 14-MAY-1999; 99US-0134137.
 PR 12-MAY-2000; 2000US-0570293.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Feild J, Yue L, Ellens H;
 XX
 DR WPI; 2001-016235/02.
 DR N-PSDB; AAC83979.
 XX
 XX Murine organic anion transporter 6 polypeptide, useful for identifying
 PT agonists/antagonists that are useful in treatment of cancer, kidney
 PT disease, autoimmune disease, inflammation and cardiovascular disease -
 XX
 PS Claim 2; Page 28-29; 32pp; English.
 XX
 CC The present sequence is murine organic anion transporter 6 (mOATP6).
 CC mOATP6 protein is useful for screening compounds which inhibit or
 CC stimulate the function of mOATP6 and also compounds that neither agonise
 CC nor antagonise OATP6. The identified agonists and antagonists are useful
 CC for prevention and treatment of human diseases, including cancer,
 CC inflammation, cardiovascular disease, central nervous system disorders,
 CC kidney diseases, liver disease and autoimmune diseases.
 XX
 SQ Sequence 540 AA;

Query Match 25.1%; Score 724.5; DB 22; Length 540;
 Best Local Similarity 34.4%; Pred. NO. 4e-70;
 Matches 187; Conservative 89; Mismatches 237; Indels 31; Gaps 13;

QY 4 YDEVTAFLGFWGFPQ--RLIFFELLSASIIIPNGFTGLSSVVELIATPEHRCRPDA-ANLS- 59
 Db 3 feellhkvvgfpgfqlrnvlialprflpmhf--llplfmaavpahcalpdpapanlsh 60
 QY 60 -SAWRNHTVPLRRDGRVPHSCRRYRLATIANFSAIGLEPGRDVDLGLQLEQ----SCL 114
 Db 61 qdlwlkthip-retdsf--ssclrfaypqalpnlvtlgtv--ynsgepegepltpcs 114
 QY 115 DGWEFSQDVLSTIVTEWNLVEDDDWKAPLITISLFFVGLGSLTSRFRGKNVLF 174
 Db 115 qgweydrsfssfiatiedwldvcqgrglnkvstctcfcfigvllgavvyeylsdrfgrirll 174

QY 175 VTMQMOTGFSFLOIFSKNFEMFVFLVFLVGMGOISNYAAAFVLGTTEILGKSVRIIFSTLG 234
 Db 175 vayvstlalglmsaasvnyimfvtttrmltgsalagfttiivlplelewidvehrtvagvis 234
 QY 235 VCIFYAFGYMVLPLFAYFTRDMRLMLVALTMPGVLVCVALMWFIPSPRWLISQGRFBEAE 294
 Db 235 t-ftwtggvllltlvgyllirswrwilllaatlpcvpgiliswwwpesarwlltqgrveeak 293
 QY 295 VIIRKAAKANGIVVPSTIPDPSELDLSSKKQ--QSHNTLDLLRTWNTIRMTIMSIMLWM 352
 Db 294 kylisacklmgprlsedsisqeaInkvitmervsgrpsyldlfrtsqirhvsicommumwf 353
 QY 353 TISVGYEGLSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLMLLQYLPRRYS-----MATAL 407
 Db 354 gvnfsyyglltdasgigtvyyqtqllfgavevpskitvffvlrvlgrlitesagmlalatal 413
 QY 408 FLGGSVLLFMQLVPPDLYLATVILVMVGKFGVTAAFSMWVYVTAELPYTVVRNMGVYSS 467
 Db 414 tfgisl-----lvssdtkswitalvvgkafseaaftaylftselypvlrqtgmfgta 468
 QY 468 TASRLGSTLSPYFVYVLCAYDRELPYILMGLSLTILTAILTLFPESEFGTLPDPTIDQMLRV 527
 Db 469 ligrlgaslaplvvllgdvwlilpklayggisflaactvlllpetkkaqlpetiqdver- 527
 QY 528 KGMM 531
 Db 528 kgmk 531

RESULT 13
 AAY51249
 ID AAY51249 standard; Protein; 535 AA.
 XX
 AC AAY51249;
 XX
 DT 14-APR-2000 (first entry)
 XX
 DE Rat liver anion transporter protein OAT2.
 XX
 KW OAT2; anion transporter; liver; rat; dicarboxylic acid; prostaglandin;
 KW non-steroid anti-inflammatory agent; anti-tumor.
 XX
 OS Rattus sp.
 XX
 PN JP11346779-A.
 XX
 PD 21-DEC-1999.
 XX
 PF 03-JUN-1998; 98JP-0169174.
 XX
 PR 03-JUN-1998; 98JP-0169174.
 XX
 PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 XX
 DR WPI; 2000-109694/10.
 DR N-PSDB; AAZ44679.
 XX
 PT A liver-specific org. anion transporter - can transport various
 PT substances including dicarboxylic acids
 XX
 PS Claim 2; Page 5-6; 13pp; Japanese.
 XX
 CC This invention describes a novel rat liver-specific org. anion
 CC transporter OAT2. The liver-specific org. anion transporter OAT2
 CC can transport various substances including dicarboxylic acids,
 CC prostaglandins, non-steroid anti-inflammatory agents and anti-tumor
 CC agents. This sequence represents the rat liver anion transporter OAT2
 CC protein described in the method of the invention.
 XX
 SQ Sequence 535 AA;

Query Match 24.9%; Score 719; DB 21; Length 535;

QY 216 VLGEILGKSVRIIFSTGLGVCIFYAF--GYMVLPLFAYFIRDRMLLVALTMPGVLCVAL 273
 Db 223 tlltefvggyr--rttailyqmaftvglvglagvayalpdrwqlavslptflily 279
 QY 274 WNFIPESPRMLISQGRFEAEVIRKAAKANGIVVPSTIFDSELDLSSKKQOSHILD 333
 Db 280 ywfypesprwllsqktrravimeqldqngkvpdpadlkmclceadasekr--spsfad 337
 QY 334 LLRTWNRMTVIMSIMLMTISVGYFGLSLDTPNLHGDFVNCFLSAMVEVPAYVLAAML 393
 Db 338 lfrtptlrkhtvilmlywfscavlygglmhvgatganlylaffysslvefpaafiliivt 397
 QY 394 LOYLPRYSMATALFLGGVLLFMQLVPPDLYLALVLMVWKGFGVTAFAFSWYVYTAEL 453
 Db 398 idrigrylpaasnlvtgaacclimifiphelwlnvtlaclgrmgativilgmvcivnael 457
 QY 454 YPTVVRNMGVGVSTASRLGSLSPYFY--LGAYDRFLPYLLMGSLTILTAILTLFLPES 512
 Db 458 yptfirnlgmvmvscaldlgiftpmvfrlmevwgalpiliifvglitagamtillipet 517
 QY 513 FGTPLPDIDQMLRVKGMKHKTPSHTRMLK 543
 Db 518 kgvalpetieaeal-grrksakentiyiq 547

RESULT 15
 AAY92902
 ID AAY92902 standard; Protein; 542 AA.
 XX
 AC AAY92902;
 XX
 DT 26-SEP-2000 (first entry)
 XX
 DE Human cerebral organic anion transporter OAT3 protein.
 XX
 KW Human; organic anion transporter protein; OAT3; cerebral tissue.
 XX
 OS Homo sapiens.
 XX
 PN WO200017237-A1.
 XX
 PD 30-MAR-2000.
 XX
 PF 20-SEP-1999; 99WO-JP05120.
 XX
 PR 18-SEP-1998; 98JP-0265126.
 XX
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 XX
 PI Endou H, Sekine T, Kusuvara H;
 XX
 DR WPI; 2000-283546/24.
 XX
 DR N-PSDB; AAA11146.
 XX

Organic anion transporter protein OAT3 expressed in cerebral tissue,
 PT antibodies to it and gene encoding it for study of cerebral transport
 PT of anions including drug molecules -
 XX
 XX Claim 2; Page 18-20; 48pp; Japanese.
 XX
 CC This sequence represents a human organic anion transporter protein OAT3.
 CC The sequence is used to regulate the transport of anions in cerebral
 CC tissue, including drug molecules, and regulation of this transport.
 XX
 SQ Sequence 542 AA;

Query Match 24.0%; Score 693; DB 21; Length 542;
 Best Local Similarity 32.9%; Pred. No. 1.1e-66;
 Matches 179; Conservative 92; Mismatches 229; Indels 44; Gaps 13;
 QY 4 YDEVTAFLGHWGPFQRIFFLLSASIPNGFTGLSSVFLIATPEHRCRVPDAANLSSAWR 63

Db 3 fseildrvsmghfqlhvailglpilmanhllqiftaatpvhrcrphnas-tgpw- 60
 QY 64 NHTVPLRLDGRGVRPHSCRRY---RLATIANFSAALGLRPRDVLGQLEQESCLDGHFEFS 120
 Db 61 --vlpwg-pngk--perclrfvhpnpnaslpndtqramep-----cidgw--- 99
 QY 121 QDVYLSV---IVTEWNLVCEDDWKAPLITSLFFVGVLLGSFISGQSLDRFGKKNVLFVTM 177
 Db 100 --vynstkdsivtcwdlvcnshklkemaqsfmagilgglvlgdlsdrfrrpiltcsy 157
 QY 178 GMOTGFSFLQIFSKNFEMFVFLVFLVGMQGISNYVAAVFLGTEILGKSVRIIFST-LGVG 236
 Db 158 lllaaagsgaafspftiyvmvfrlclgfgisgitlsvilnvwvptrmraimstalgyc 217
 QY 237 IFYAFGVWVLPFAYFIRDRMLLVALTMPGVLCVALMWFIPESPRMLISQGRFEAEVI 296
 Db 218 --ytfqgfilpglayalpqrwqltvsipffvflsswvtpesirwlvsgksealki 275
 QY 297 IRKAAKANGIVVPSTIFDSELDLSSKKQOS-----HNILDLRTWNRMTVIMSIMLW 351
 Db 276 lrrvavfngkkeeerlsleek-lnlqkeislakakytasdlfrimprrmtfclslaw 334
 QY 352 MTISVGYFGLSLDTPNLHGDFVNCFLSAMVEVPAYVLAAMLLOYLPRYSMATALFLGG 411
 Db 335 fatgfayyslamgveefgvnlyilqilfggvdpakfifilsyigrhttqaaallag 394
 QY 412 SVLLFMQLVPPDLYLALVLMVWKGFGVTAFAFSWYVYTAELPYTVVRNMGVGVSTASR 471
 Db 395 gailaltfvpidlqtrvtvrlavfvgkclsssfcliytselyptvirqtgmvgvsnlwr 454
 QY 472 LGSILSPYFVYVIGAYDRFLPYLLMGSLTILTAILTLFLPESFGTLPDIDQM---LRV 527
 Db 455 vgsmsvplvkitgevqfipniygitalliggsaalflpetlloqplpetiedlenwsra 514
 QY 528 KGMK 531
 Db 515 kkp 518

Search completed: August 16, 2001, 13:58:26
 Job time: 208 sec

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:58:52 ; Search time 20.69 Seconds
(without alignments)
554.317 Million cell updates/sec

Title: US-09-521-195-3
Perfect score: 2883
Sequence: 1 MRDYDVTAFLGEGWGFQRL.....HTRMLKDGQERTILKSTAF 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|------------------|
| 1 | 726 | 25.2 | 555 | 3 | US-08-501-572-3 |
| 2 | 726 | 25.2 | 555 | 3 | US-09-040-444-3 |
| 3 | 718 | 24.9 | 556 | 3 | US-08-501-572-1 |
| 4 | 718 | 24.9 | 556 | 3 | US-09-040-444-1 |
| 5 | 714.5 | 24.8 | 553 | 3 | US-08-501-572-2 |
| 6 | 714.5 | 24.8 | 553 | 3 | US-09-040-444-2 |
| 7 | 652.5 | 22.6 | 537 | 2 | US-08-647-397-2 |
| 8 | 234 | 8.1 | 494 | 2 | US-09-031-392-5 |
| 9 | 234 | 8.1 | 494 | 2 | US-09-299-549-5 |
| 10 | 208 | 7.2 | 492 | 2 | US-08-355-844-3 |
| 11 | 208 | 7.2 | 492 | 5 | PCT-US95-16126-3 |
| 12 | 203 | 7.0 | 493 | 2 | US-09-031-392-10 |
| 13 | 203 | 7.0 | 493 | 4 | US-09-299-549-10 |
| 14 | 198 | 6.9 | 109 | 2 | US-08-647-397-4 |
| 15 | 193 | 6.7 | 509 | 2 | US-09-031-392-6 |
| 16 | 193 | 6.7 | 509 | 4 | US-09-299-549-6 |
| 17 | 186 | 6.5 | 500 | 2 | US-09-031-392-7 |
| 18 | 186 | 6.5 | 500 | 4 | US-09-299-549-7 |
| 19 | 184 | 6.4 | 563 | 2 | US-09-031-392-2 |
| 20 | 184 | 6.4 | 563 | 4 | US-09-299-549-2 |
| 21 | 175 | 6.1 | 488 | 2 | US-08-928-692-11 |
| 22 | 168.5 | 5.8 | 534 | 2 | US-09-031-392-4 |
| 23 | 168.5 | 5.8 | 534 | 4 | US-09-299-549-4 |
| 24 | 155 | 5.4 | 383 | 2 | US-09-031-392-3 |
| 25 | 155 | 5.4 | 383 | 4 | US-09-299-549-3 |
| 26 | 153.5 | 5.3 | 524 | 2 | US-08-928-692-12 |
| 27 | 149 | 5.2 | 584 | 2 | US-08-928-692-13 |

| | | | | | |
|----|-------|-----|------|---|-------------------|
| 28 | 131 | 4.5 | 488 | 2 | US-08-928-692-10 |
| 29 | 110 | 3.8 | 473 | 1 | US-08-597-236-13 |
| 30 | 110 | 3.8 | 473 | 1 | US-08-746-682A-13 |
| 31 | 104.5 | 3.6 | 1835 | 3 | US-08-836-325-15 |
| 32 | 101.5 | 3.5 | 341 | 1 | US-08-423-564-5 |
| 33 | 97 | 3.4 | 342 | 3 | US-08-785-928-1 |
| 34 | 97 | 3.4 | 342 | 3 | US-08-728-603-17 |
| 35 | 97 | 3.4 | 916 | 2 | US-08-928-692-58 |
| 36 | 94.5 | 3.3 | 390 | 5 | PCT-US92-02091-8 |
| 37 | 93.5 | 3.2 | 467 | 2 | US-08-805-118-3 |
| 38 | 93.5 | 3.2 | 480 | 2 | US-08-724-394A-9 |
| 39 | 93 | 3.2 | 560 | 2 | US-08-805-118-4 |
| 40 | 92.5 | 3.2 | 419 | 1 | US-08-385-186-2 |
| 41 | 92.5 | 3.2 | 568 | 4 | US-08-637-823B-27 |
| 42 | 92.5 | 3.2 | 584 | 4 | US-08-637-823B-28 |
| 43 | 92 | 3.2 | 429 | 2 | US-08-677-049-5 |
| 44 | 92 | 3.2 | 503 | 4 | US-09-068-195-24 |
| 45 | 92 | 3.2 | 548 | 3 | US-08-903-139B-28 |

ALIGNMENTS

RESULT 1
US-08-501-572-3
; Sequence 3, Application US/08501572
; Patent No. 6063623

; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin:Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-501-572-3

Query Match 25.2%; Score 726; DB 3; Length 555;
Best Local Similarity 35.0%; Pred. No. 5.2e-68;
Matches 187; Conservative 98; Mismatches 223; Indels 26; Gaps 14;
QY 5 DEVTAFLGEGWGFQRLIFF---LISASITPNGTGLSSVFLIATPEHRCRVPDANLUS-- 59

```

;
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-040-444-3

Query Match      25.2%; Score 726; DB 3; Length 555;
Best Local Similarity 35.0%; Pred. No. 5.2e-68;
Matches 187; Conservative 98; Mismatches 223; Indels 26; Gaps 14;

QY 5  DEVTAFLGGEWPFQRIFF---LLSASIPNGFTGLSVFIATPEHRCRVPDAANLS-- 59
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 6  DDVLEHGGEFFHFKQMEFFLLALLSATFAP1-YVGI--VFLGFTPDHRCRSPGVAELSRL 62
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 60 SAWR-----NHTVPLRLRDGREVPHSCRRYRLA-TIANESALGLEPGRDVDLQLEQESC 113
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 63 COWSPAELNVTVPGPAGEASPRQCRRYEVDWVNQSTFDCVDPLASLDTNRSRLPLGPC 122
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 114 LDGWEFSQDVLSTIVTWNLVCEDDWKAPLTISLFPVGVLLGSFISGQLSDRGRKNVL 173
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 123 RDGWY--ETPGSSIVTEFNLCANSMWLDLQSSVNVGFFIGSMISGYIADREFGRKLC 180
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 174 FVTGMQTFGSFLOIFSKNFEMFVFLVFLVGMGOISNYAAAFVLGTEILGKSVRIESTL 233
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 181 LTTVLINAAAGVLAISPTYTWMILIFRLIOGLVSKAGWLGIIILITEFVGGRYR---RTV 237
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 234 GVC--IFYAFGYMVLPAFYFIRDMRMLLVALTMPGVLCVAMWFIPESPRWLISQGRFE 291
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 238 GIFQVAYTVGLLVAGVAYALPHWRWLQFTVALPNEFFLLYWCIPESPRWLISQNKNA 297
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 292 EAEVIRKAANGVIVPSTIFDPSELDQLSSKKQSHNLDLRTWNRMVMTIMSLW 351
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 298 EAMRIKHIKAKNGKSLPASL-QRLLEEETGKK-LNPSFLDIVRTPQIRKHTMILMYNW 355
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 352 MTISVGVGFLSLDTPNLHGD-IFVNCFLSAMVEVPAYVLAWLLQLYLPRRYSMATALFLG 410
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 356 FTSSVLYQGLIMHM-GLAGDNIYLDFFYSALVEFPAAFMILIIDRIGRRYPWAAASNMYA 414
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 411 GSVLLFMQLVPPDLYLYLATVLMVKGKGVTAASWVYVYTAELYPTVVRNMGVGSSTAS 470
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 415 GAACLASVFIPGDLQWLKIIISCLGRMGITMAYEIVCLVNAELYPTFIRNLGVHCSSMC 474
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 471 RLGSILSPFYVY-IGAYDRFLPYLMGSLTILTAITLFLPESFGPLPDTIDQ 523
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 475 DIGGIITPFLVYRLTNWLEPLMWVGVGLVAGGLALLPETKALPETIEE 528
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 3
US-09-040-444-3
; Sequence 3, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; NUMBER OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Parabow,Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```

;
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-040-444-3

Query Match      25.2%; Score 726; DB 3; Length 555;
Best Local Similarity 35.0%; Pred. No. 5.2e-68;
Matches 187; Conservative 98; Mismatches 223; Indels 26; Gaps 14;

QY 5  DEVTAFLGGEWPFQRIFF---LLSASIPNGFTGLSVFIATPEHRCRVPDAANLS-- 59
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 6  DDVLEHGGEFFHFKQMEFFLLALLSATFAP1-YVGI--VFLGFTPDHRCRSPGVAELSRL 62
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 60 SAWR-----NHTVPLRLRDGREVPHSCRRYRLA-TIANESALGLEPGRDVDLQLEQESC 113
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 63 COWSPAELNVTVPGPAGEASPRQCRRYEVDWVNQSTFDCVDPLASLDTNRSRLPLGPC 122
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 114 LDGWEFSQDVLSTIVTWNLVCEDDWKAPLTISLFPVGVLLGSFISGQLSDRGRKNVL 173
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 123 RDGWY--ETPGSSIVTEFNLCANSMWLDLQSSVNVGFFIGSMISGYIADREFGRKLC 180
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 174 FVTGMQTFGSFLOIFSKNFEMFVFLVFLVGMGOISNYAAAFVLGTEILGKSVRIESTL 233
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 181 LTTVLINAAAGVLAISPTYTWMILIFRLIOGLVSKAGWLGIIILITEFVGGRYR---RTV 237
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 234 GVC--IFYAFGYMVLPAFYFIRDMRMLLVALTMPGVLCVAMWFIPESPRWLISQGRFE 291
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 238 GIFQVAYTVGLLVAGVAYALPHWRWLQFTVALPNEFFLLYWCIPESPRWLISQNKNA 297
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 292 EAEVIRKAANGVIVPSTIFDPSELDQLSSKKQSHNLDLRTWNRMVMTIMSLW 351
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 298 EAMRIKHIKAKNGKSLPASL-QRLLEEETGKK-LNPSFLDIVRTPQIRKHTMILMYNW 355
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 352 MTISVGVGFLSLDTPNLHGD-IFVNCFLSAMVEVPAYVLAWLLQLYLPRRYSMATALFLG 410
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 356 FTSSVLYQGLIMHM-GLAGDNIYLDFFYSALVEFPAAFMILIIDRIGRRYPWAAASNMYA 414
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 411 GSVLLFMQLVPPDLYLYLATVLMVKGKGVTAASWVYVYTAELYPTVVRNMGVGSSTAS 470
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 415 GAACLASVFIPGDLQWLKIIISCLGRMGITMAYEIVCLVNAELYPTFIRNLGVHCSSMC 474
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 471 RLGSILSPFYVY-IGAYDRFLPYLMGSLTILTAITLFLPESFGPLPDTIDQ 523
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 475 DIGGIITPFLVYRLTNWLEPLMWVGVGLVAGGLALLPETKALPETIEE 528
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 3
US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Parabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohy, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-501-572-1

```

```

Query Match 24.9%; Score 718; DB 3; Length 556;
Best Local Similarity 34.0%; Pred. No. 3.6e-67;
Matches 194; Conservative 96; Mismatches 229; Indels 52; Gaps 14;

```

```

QY 1 MRDYDEVTAFLEGEWPFQRLIFELL--SASIIPIGNGFTGLSSVFLIATPEHRCRVPDAAN 57
Db 1 MPTVDVLEQVGEFGWQKQAFLLCLISASLAPI-YVGI--VFLGFTPGHYCONPGVAE 57

QY 58 LSS--AWR-----NHTVP-LRLRDGREVPHSCRRYL-----ATIANFSAL 95
Db 58 LSQRCGWSQAEEELNYTPGLGPSDEASFLSQCMRYEVDWQNSTLDCVDP LSSLVANRSQL 117

QY 96 GLEPGRDVLDGQESCLDGWFSQDYVLSITVTEWNLVCEDDWKAPLITSLFFVGVLL 155
Db 118 PLGP-----CEHGWY--DTPGSSIVTEFNLVCGDAWKVDLFQSCVNLGFFL 162

QY 156 GSPISGQLSDRFGKKNLVFTMGOTGFSLQIFSKNPFVFLVLCMGQISNVAFAF 215
Db 163 GSVLVGYIADRGKRLCLLVTSVSGVLTA VADPTSMILLFRLQGMVSKGWSVSGY 222

QY 216 VLGTILGKSVRIIFSTLGVCFYAF--GYMVLPLFAFYFIRDRMRLVLTMPGVLCVAL 273
Db 223 TLITEFVSGYR---RTTALYQMAFTVGLVGLAGVAYAIPTDRLWQLAVSLPTFLFLY 279

QY 274 WFIPEPRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSLQDLSKKQKQSHNILD 333
Db 280 YWFPESPRWLLSQKRTTRAVRIMEQIAQKNGKVPADLKLMLCLEEDASEKR--SPSEAD 337

QY 334 LLRTWNRMTVIMSIMLWMTISVGYFGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWL 393
Db 338 LFTPNLRKHTVILMTLWFSQVGLVGLAGVAYAIPTDRLWQLAVSLPTFLFLY 397

QY 394 LQYLPYRYSMATLFLGGSVLLFMQLVPPDLYLATVLMVWGKFGVTAAFSMVYVTAEL 453
Db 398 IDRIGRIYPIAASNLYTGAACLLMIFIPHELHNLVLTACLRMGATIVLEWCLVNAEL 457

QY 454 YPTVNRMGVGSSTASRLSILSPFYV-IGAYDRFLPYILMGSLTITLTAITLFLPES 512
Db 458 YPTFIRNLGMVCSALCDLGGITFPMVPRMLVEWQALPLFLGLVGLTAGAMTLLLPET 517

QY 513 FGTPLPDITDMLRVKGMKHKRTPSPTRMLK 543
Db 518 KGVALPETIEEABL--GRRKSKAKENTIIYLQ 547

```

```

RESULT 4
US-09-040-444-1
; Sequence 1, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk

```

```

; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport of cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-040-444-1

```

```

Query Match 24.9%; Score 718; DB 3; Length 556;
Best Local Similarity 34.0%; Pred. No. 3.6e-67;
Matches 194; Conservative 96; Mismatches 229; Indels 52; Gaps 14;

```

```

QY 1 MRDYDEVTAFLEGEWPFQRLIFELL--SASIIPIGNGFTGLSSVFLIATPEHRCRVPDAAN 57
Db 1 MPTVDVLEQVGEFGWQKQAFLLCLISASLAPI-YVGI--VFLGFTPGHYCONPGVAE 57

QY 58 LSS--AWR-----NHTVP-LRLRDGREVPHSCRRYL-----ATIANFSAL 95
Db 58 LSQRCGWSQAEEELNYTPGLGPSDEASFLSQCMRYEVDWQNSTLDCVDP LSSLVANRSQL 117

QY 96 GLEPGRDVLDGQESCLDGWFSQDYVLSITVTEWNLVCEDDWKAPLITSLFFVGVLL 155
Db 118 PLGP-----CEHGWY--DTPGSSIVTEFNLVCGDAWKVDLFQSCVNLGFFL 162

QY 156 GSPISGQLSDRFGKKNLVFTMGOTGFSLQIFSKNPFVFLVLCMGQISNVAFAF 215
Db 163 GSVLVGYIADRGKRLCLLVTSVSGVLTA VADPTSMILLFRLQGMVSKGWSVSGY 222

QY 216 VLGTILGKSVRIIFSTLGVCFYAF--GYMVLPLFAFYFIRDRMRLVLTMPGVLCVAL 273
Db 223 TLITEFVSGYR---RTTALYQMAFTVGLVGLAGVAYAIPTDRLWQLAVSLPTFLFLY 279

QY 274 WFIPEPRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSLQDLSKKQKQSHNILD 333
Db 280 YWFPESPRWLLSQKRTTRAVRIMEQIAQKNGKVPADLKLMLCLEEDASEKR--SPSEAD 337

QY 334 LLRTWNRMTVIMSIMLWMTISVGYFGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWL 393
Db 338 LFTPNLRKHTVILMTLWFSQVGLVGLAGVAYAIPTDRLWQLAVSLPTFLFLY 397

QY 394 LQYLPYRYSMATLFLGGSVLLFMQLVPPDLYLATVLMVWGKFGVTAAFSMVYVTAEL 453
Db 398 IDRIGRIYPIAASNLYTGAACLLMIFIPHELHNLVLTACLRMGATIVLEWCLVNAEL 457

```

QY 454 YPTVVRNMGVGVSTASRLGSLSPYFY-IGAYDRFLPYILMGSLTILTAITLFLPES 512
 Db. 458 YPTFIRNLGMVCSALCDLGGIFTPEMFRLEMEVQALPLILFVGLGTAGAMTILLPET 517
 QY 513 FGTPLPDTIDQMLRVKGMKHKRTPSHRMLK 543
 Db 518 KGVALPETIEEAEHL-GRKRSKAKENTYLO 547

RESULT 5
 US-08-501-572-2
 ; Sequence 2, Application US/08501572
 ; Patent No. 6063623
 ; GENERAL INFORMATION:
 ; APPLICANT: Koepsell, Hermann
 ; APPLICANT: Grundeman, Dirk
 ; APPLICANT: Gorboulev, Valentin
 ; TITLE OF INVENTION: Transport protein which effects the
 ; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
 ; TITLE OF INVENTION: DNA sequences encoding it and their use.
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
 ; STREET: 1300 I Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/501,572
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Toohey, Kimberlin M
 ; REGISTRATION NUMBER: 35,391
 ; REFERENCE/DOCKET NUMBER: 02481.1453-00000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)408-4000
 ; TELEFAX: (202)408-4400
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 553 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-501-572-2

Query Match 24.8%; Score 714.5; DB 3; Length 553;
 Best Local Similarity 35.4%; Pred. No. 8.4e-67;
 Matches 194; Conservative 87; Mismatches 216; Indels 51; Gaps 13;

QY 1 MRDYDEVTAFLCEWGFQF---LIFELLASITIPNGFTGLSSVFLIATPEHRCRVPAAN 57
 Db 1 MPTVDDILEQVGESWFOKQAFILICLLSAFAP---ICVGIVFLGFTPDHHCQSPGVAE 57
 QY 58 LSS---AWR-----NHTVPLRLRDGVEPHSCRRYR-----LATIA-NFSALG 96
 Db 58 LSQRGWSPAEELNVTVPGLGPAEAFGLQCRREVDNQSALSCVDPPLASLATNRSLP 117
 QY 97 LEFGROVDLQLEQSCLDGWEFFSDQVYLSIVTEWNLVCEDDWKAPLTISLFFVGVLLG 156
 Db 118 LGP-----DQGGWY---DTGSSIVTEFNLVCADSKWLDLFSQCLNAGFFFG 162
 QY 157 SPISGLSDRFGRKKNVLFVTWGMQTFGSLQIFSKNFEMFVFLVGMGQISNVAAFV 216

Db 163 SLGVGVFADRFGRKLCCLLGTVLVNAVSGVLMAFSPNYSMLLRLLOGLVSKGNWAGYT 222
 QY 217 LGTEILGKSVRIIFSTLGVCFYAF--GYMVLPLFAFYFIRDMRLMLVALTMPGVLCAVLM 274
 Db 223 LIEFVGSGSR--RTVAIMYQMAFTVGLVALTGLAYALPHWRWLQVLAVALSLPFLFLLY 279
 QY 275 WFPESPRWLISQGRFEAEVIRKAANGIVVVPSTIFDPSLQDLSSKKQOOSHNLDL 334
 Db 280 WCVPESPRWLLSQKRNTAIKIMDHIAQKNGKLPADLKLMSLEEDVTEK--LSPSFADL 337
 QY 335 LRTWNIRMVTIMSLMWTISVGVFGLSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLLL 394
 Db 338 FRTPRLKRRTFILMYLWFTDSVLYQGLIHMGATSNLYDLFDLYSALVEIPGAFIALITI 397
 QY 395 QYLPRRYSMATALFLGSGVLLFMQLVPPDLYLATVLVMVKGFGVTAAFSMVVYTAELY 454
 Db 398 DRVGRIYPMVAVSNLLAGAACLVI-FISPDHLHLNLIIMCVGRMGITIAIOMICLVNAELY 456
 QY 455 PTVVRNMGVGVSTASRLGSLSPYFY-IGAYDRFLPYILMGSLTILTAITLFLPESF 513
 Db 457 PTFVRLNRVMVCSALCDLGGIFTPEFVFRLEVMQALPLILFAVLGLAAGVTLILLPETK 516
 QY 514 GTPLPDTI 521
 Db 517 GDALPETM 524

RESULT 6
 US-09-040-444-2
 ; Sequence 2, Application US/09040444
 ; Patent No. 6063766
 ; GENERAL INFORMATION:
 ; APPLICANT: Koepsell, Hermann
 ; APPLICANT: Grundeman, Dirk
 ; APPLICANT: Gorboulev, Valentin
 ; TITLE OF INVENTION: Transport protein which effects the
 ; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
 ; TITLE OF INVENTION: DNA sequences encoding it and their use.
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
 ; STREET: 1300 I Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/040,444
 ; FILING DATE: March 18, 1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: O'Connor, Steven P
 ; REGISTRATION NUMBER: 41,225
 ; REFERENCE/DOCKET NUMBER: 2481.1453-01
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)408-4000
 ; TELEFAX: (202)408-4400
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 553 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-09-040-444-2

Query Match 24.8%; Score 714.5; DB 3; Length 553;

Best Local Similarity 35.4%; Pred. No. 8.4e-67; Matches 194; Conservative 87; Mismatches 216; Indels 51; Gaps 13;

QY 1 MRDVEYTAFLGEGPQR---LFFELLSASIPNGFTGLSSVFLIATPEHRCRVDPDAAN 57
Db 1 MPTVDDILEQVSGWFKQAFLLCLLSAFAF---ICVGVPLGFTPDHHCQSPQVAE 57
QY 58 LSS--AWR-----NHTVPLRLRDGREVPHSCRRY-----LATIA-NFSALG 96
Db 58 LSQRCGNSPAEELNVTPLGSPAGEAFGLQCRRYEVDMNQSALESCVDPPLASLATNRSLP 117
QY 97 LEPRDVLGQLEBESCLDGEFSDQVYLSITVTEWNLVCEDDWKAPLTIISLFPVGLLG 156
Db 118 LGP-----CODGWY--DTPGSIIVTEFNLVCAWSKLDLFOCLNAGFFFG 162
QY 157 SFISQSLDRGRKNVFLVMTGMOTGFSFLOIFSKNFEMFVFLVGMGOISNYAAVF 216
Db 163 SLGVGFADREGKLCILGTVLVNAVSVLMAFSPNYMSMLLFRLLQGLYSKGNMAGYT 222
QY 217 LGTEILGKSVRIIESTLGCIFYAF--GYMVLPLFAFYRDRWMLLVALTMGVLVALW 274
Db 223 LITEFVSGSR---RTVAIMYQMAFTGLVALTGLAYALPHRWLQLAVALPTFLFLYY 279
QY 275 WFIPESRWLLSOGREFAEVLIRKAANGIVVPSTIFDPSLQDLSSKKQOQSHNILD 334
Db 280 MCVPEPRLWLLSOKRTEAIIKIMDHIAOKNGKLPADKMLSLDEEDVTEK--LSPSFADL 337
QY 335 LRTNIRMTIMSIMLMTISVGVFGLSLDTPNLHGDIYFNCFLSAMVEVPAYVLAWL 394
Db 338 FRTPLRKRTFTLMLWFTDSVLYQGLILHMGATSGNLYLDLYLSALVEIPGAFIALITI 397
QY 395 QYLPYRYSMATLFLGGSVLLFMQLVPPDLYLATVLMVKGKFGVTAFAFSVYVTAELY 454
Db 398 DRVGRIYMAVSNLLAGAACLV-FISPDHLWLNIIIMCVGRMGITIAIOMICLVNAELY 456
QY 455 PTVVRNMGVGVSTASRLSGSILSPYFV-LGAYDRFLPYILMGSILTITLITFLPESF 513
Db 457 PTFVRLNLRVWVSSICDGGIITPFIIVFLREVWQALPILFAVLGLLAAGVTLILLPETK 516
QY 514 GTPLPDTI 521
Db 517 GDALPETM 524

RESULT 7
US-08-647-397-2
; Sequence 2, Application US/08647397
; Patent No. 5972702
; GENERAL INFORMATION:
; APPLICANT: Beier, David R.
; TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,397
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7048

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-647-397-2

Query Match 22.6%; Score 652.5; DB 2; Length 537;
Best Local Similarity 29.9%; Pred. No. 2.8e-60; Indels 41; Gaps 12;
Matches 167; Conservative 104; Mismatches 246;

QY 4 YDEVTAFLGEMGPPQRLIFFELLSASIPNGFTGLSSVFLIATPEHRCRVDPDAANLSSAWR 63
Db 3 FSEILDRVSGMGFPQYLHVTLTLLALPILGIANHNLQIFATTATPDHHCRRPPNASL-EPW- 60
QY 64 NHTVPLRLRDGREVPHSCRRY---RLATIANFSALGLEPRDVLGQLEBESCLDGEFES 120
Db 61 --VPLG-PNGK--PEKCLRFVHLNPNASLPNDTQGATEP-----CLDGIWYN 102
QY 121 QDVYLSITVTEWNLVCEDDWKAPLTIISLFFVGLLGSFISGQSLDRGRKNVFLVTMGMO 180
Db 103 ST--RTIVTEWDLVCGSNKLEMAQSVFMAGILVGGPVFGLSDRFGKRPILTWSYLL 160
QY 181 TGFSEFLQIFSKNFEMFVFLVGMGOISNYAAVFVLTGTEILGKSVRIIFS-TLGVCIY 239
Db 161 AASGSSAAFSPLTVVYIMIFRFLCGCSISGISLSTIILNVWVPTSTRAISSTTIGYC--Y 218
QY 240 AFGWVLPFAFYRDRWMLLVALTMGVLCAVLMWFIPESRWLLSOGREFAEVLIRK 299
Db 219 TIGQFVLGLAYAVPQWRQLQSVSAFFIFSLLSMWPESIRWLVLSKFSRALKTLOR 278
QY 300 AARANGIVVPSTIFDPSL-----QDLSSKKQOQSHNILDRLTNIRMTIMSIMLMTI 354
Db 279 VAFNGKKEGEKLTVEELKFNQKDIKSAVK-YGLSDLFVSLIRRVYTFCLSLAWFAT 337
QY 355 SVGVFGLSLDTPNLHGDIYFNCFLSAMVEVPAYVLAWLLOYLPRRYSMATLFLGGSVL 414
Db 338 GFAYISLAMGVVEEFGVNIYILQIFGGVDIPAKFITILSISYLGRRITQGLLILAGVAI 397
QY 415 LFMQVPPDLYLATVLMVKGKFGVTAFAFSVYVTAELYPTVVRNMGVGVSTASRLGS 474
Db 398 LALIFVSSEMQLDRALAVFGKCLSGSFCLFYTSSELYPTVLRQTGMGINSINWARGS 457
QY 475 ILSPYFVVLGAYDRFLPYILMGSILTITLITFLPESFGTLPDITID-----QMLRV 527
Db 458 MIAPLVKITGELQFPFIPNVIFWTMLLGGSAAPFLLETNLNRPLETIEDIQDWYQOTKKT 517
QY 528 KGMKHKRTPSHRMLKOG 545
Db 518 KQPEAEKASQTIPLKTG 535

RESULT 8
US-09-031-392-5
; Sequence 5, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA

ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/031,392
 FILING DATE: 26-FEB-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Meiklejohn, Ph.D., Anita L.
 REGISTRATION NUMBER: 35,283
 REFERENCE/DOCKET NUMBER: 07334/072001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 494 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-031-392-5

Query Match 8.1%; Score 234; DB 2; Length 494;
 Best Local Similarity 24.7%; Pred. No. 3.6e-16;
 Matches 112; Conservative 85; Mismatches 158; Indels 98; Gaps 23;

```

QY 144 L T S L P F V G V L L G S F T S G L S D R E G R K ----- N V L F V T M G M Q T G F S F L Q I F S K N F E M F V 197
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 65 L S V A I F S V G G M I G S F S G L F V N R F G R N S M L I V N L L A I A G G C L M G F C K I --- A E S V E M L I 121

QY 198 V L F V L V G M -- G Q I S N Y V A A F V L G T E I L G K S V R I F S T L ----- G V C I F Y A F G Y M V L P L 248
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 122 L G R L I I G L F C L G T G F V P M Y I -- G E I S P T A L R G A F G T L N Q L G I V I G L V A Q I F G L K V I -- 177

QY 249 F A Y F I R D - W R M L I V A L T M P G V L C V A L M W F I P E S P R W - L I S O G R F E E A E V I I R K A A K A N G I 306
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 178 -- L G T E D L M P L L G T I L P A I I O C A A L P C P E S P R E L L I N R K E E K A K E I L Q R L M G T E D V 235

QY 307 V V P S T I F D P S E L Q D L S K K Q O S H - N I L D L L R T W N I R M V T I M S I M L W M T ----- I S V G Y 358
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 236 A Q ----- D I Q E M K D E S W R M S Q E K Q V T V L E F R A P N Y R Q P I I I S I M L Q S Q L S G I N A V F Y 290

QY 359 F G L S L D P T P N L H G D I F V N C F L S A M V E V P A Y ----- V L A M L L Q Y L P R R Y S M A T 405
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 291 Y S T G I ----- F K D A G V Q E P V A T I G A V V V N T I F T V V S V F L V E R A G R R -- T L H 335

QY 406 A L F L G G --- S V L L F M Q L V P P D L Y ----- Y L A T V L V M V G K F G V T A A F S M V Y V Y T A E L Y P 455
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 336 L I G L G M A F C S I L M T I S L L K D N Y S M S F I C I G A I L V F A F F E I G P G - P I P W F I V A E L F G 394

QY 456 T V V R N M G V G S S --- T A S R L G S I L S P Y F V - Y L G A Y D R F L P Y I L M G S L T I L T A I L T L F - L 509
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 395 Q G R P R A M A V A C S N W T S N F L V G L L F P S A T F Y L G A Y --- V F I V F V F L V F W F T F F K V 450

QY 510 P E S F G T P L P D T I D O M L R V ----- K G M K H R K T P 536
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 451 P E T R G R --- T F E I T R A F E G V Q V T G T R G E K G P 479
  
```

RESULT 9
 US-09-299-549-5
 Sequence 5, Application US/09299549
 Patent No. 6136547
 GENERAL INFORMATION:
 APPLICANT: Tartaglia, Louis A.
 APPLICANT: Weng, Xun
 TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
 TITLE OF INVENTION: GLUTEX AND USES THEREOF
 NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/299,549
 FILING DATE: 26-APR-1999
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/031,392
 FILING DATE: 26-FEB-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Meiklejohn, Ph.D., Anita L.
 REGISTRATION NUMBER: 35,283
 REFERENCE/DOCKET NUMBER: 07334/072002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 494 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-299-549-5

Query Match 8.1%; Score 234; DB 4; Length 494;
 Best Local Similarity 24.7%; Pred. No. 3.6e-16;
 Matches 112; Conservative 85; Mismatches 158; Indels 98; Gaps 23;

```

QY 144 L T S L P F V G V L L G S F T S G L S D R E G R K ----- N V L F V T M G M Q T G F S F L Q I F S K N F E M F V 197
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 65 L S V A I F S V G G M I G S F S G L F V N R F G R N S M L I V N L L A I A G G C L M G F C K I --- A E S V E M L I 121

QY 198 V L F V L V G M -- G Q I S N Y V A A F V L G T E I L G K S V R I F S T L ----- G V C I F Y A F G Y M V L P L 248
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 122 L G R L I I G L F C L G T G F V P M Y I -- G E I S P T A L R G A F G T L N Q L G I V I G L V A Q I F G L K V I -- 177

QY 249 F A Y F I R D - W R M L I V A L T M P G V L C V A L M W F I P E S P R W - L I S O G R F E E A E V I I R K A A K A N G I 306
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 178 -- L G T E D L M P L L G T I L P A I I O C A A L P C P E S P R E L L I N R K E E K A K E I L Q R L M G T E D V 235

QY 307 V V P S T I F D P S E L Q D L S K K Q O S H - N I L D L L R T W N I R M V T I M S I M L W M T ----- I S V G Y 358
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 236 A Q ----- D I Q E M K D E S W R M S Q E K Q V T V L E F R A P N Y R Q P I I I S I M L Q S Q L S G I N A V F Y 290

QY 359 F G L S L D P T P N L H G D I F V N C F L S A M V E V P A Y ----- V L A M L L Q Y L P R R Y S M A T 405
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 291 Y S T G I ----- F K D A G V Q E P V A T I G A V V V N T I F T V V S V F L V E R A G R R -- T L H 335

QY 406 A L F L G G --- S V L L F M Q L V P P D L Y ----- Y L A T V L V M V G K F G V T A A F S M V Y V Y T A E L Y P 455
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 336 L I G L G M A F C S I L M T I S L L K D N Y S M S F I C I G A I L V F A F F E I G P G - P I P W F I V A E L F G 394

QY 456 T V V R N M G V G S S --- T A S R L G S I L S P Y F V - Y L G A Y D R F L P Y I L M G S L T I L T A I L T L F - L 509
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 395 Q G R P R A M A V A C S N W T S N F L V G L L F P S A T F Y L G A Y --- V F I V F V F L V F W F T F F K V 450

QY 510 P E S F G T P L P D T I D O M L R V ----- K G M K H R K T P 536
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 451 P E T R G R --- T F E I T R A F E G V Q V T G T R G E K G P 479
  
```

RESULT 10

US-08-355-844-3

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

| Query Match | 7.2% | Score 208; | DB 2; | Length 492; |
|-----------------------|--------------|---|-----------------|----------------------|
| Best Local Similarity | 22.2%; | Pred. No. 2e-13; | | |
| Matches 108; | Conservative | 87; | Mismatches 165; | Indels 126; Gaps 22; |
| Qy | 119 | FSQDYLSTIWTENWLVCEDDWKAPLTSLTFEFGVLLGSFISGQLSDRGRK-----NV | 172 | |
| Db | 52 | YGESILPTTLTNS-----LSVAIFSVGGMIGSFVGLFVNRGRNSLMNML | 101 | |
| Qy | 173 | LFVTMGQTGSPFLQIFSKNPFVFLVGM--GQISNVYAAFLVTEILGKSVRIIF | 230 | |
| Db | 102 | LAFYSAVILMGFSKL--GKSFEMTLFLFIIGVYCGLTGFVPYV--GEVSPFAFGAL | 156 | |
| Qy | 231 | STL-----GVCIFYAFGYVVLPLFAFYIRD-----WRMLIVALTMPGVLCVALMW | 275 | |
| Db | 157 | GTLHOLGIVVGULLTAQVGL-----DSIMGKDLWPLLLLSIIFIPALLOQIVIP | 205 | |
| Qy | 276 | FIPSPRW-LISQGFEBEAIVIRKAAKANGIVVPSTTFDFSELODLSKKQOOSH----- | 329 | |
| Db | 206 | FCPSEPRELLNRNDEENRAKSVLKLRTGADVT-----HDLQEMKEESQMMRE | 254 | |
| Qy | 330 | ---NILDLLRTWNIRAWTIVMSITLWMT-----ISVGYFGLSLDPPNLHGDFVNCFLSA | 380 | |

[illegible]

```

RESULT 11
PCT-US95-16126-3
; Sequence 3, Application PC/TUS9516126
; GENERAL INFORMATION:
; APPLICANT: Fischbarg, Jorge
; APPLICANT: Czegledy, Ferenc
; APPLICANT: Iserovich, Pavel
; APPLICANT: Li, Jun
; APPLICANT: Cheung, Min
; TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
; TITLE OF INVENTION: STRUCTURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/355,844
; FILING DATE: 14-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Tang, Henry Y.S.
; REGISTRATION NUMBER: 29,705
; REFERENCE/DOCKET NUMBER: A29927-50/29910
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2586
; TELEFAX: 212-765-2519
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Human
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..492
; OTHER INFORMATION: Facilitative glucose transport
; OTHER INFORMATION: Glut1 protein
PCT-US95-16126-3

```

```

Query Match          7.2%  Score 208; DB 5; Length 492;
Best Local Similarity 22.2%  Pred No. 2e-13;
Matches 108; Conservative 87; Mismatches 165; Indels 126; Gaps 22;

QY 119 FSDVYLSTIVTWNIVCEDDKAPITISLFFVGVLLGSFISQGLSDRGRK-----NV 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52 YGESILPTTLTLLNS-----LSVAIESVGGMIGSFVGLFVNREGRRNMLMNL 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 LFTVMGOTGFSFLOIFSKNFEMFVFLVGM--COISNYVAAFVLGTGLGKSVRIIF 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 LAFVSAVLGMGFSK-----GKSFEMLIILGRFIIGYCGLTATGFPVMYV--GEVSPATFRGAL 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 STL-----GVCIFYAFGYMVLPLFAYEIRD-----WRMLLVALTMPGVLCVALMW 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 GTLHQLGIIVGILIAQVFL-----DSIMGNKDLWPLLSIIFIPALLOCIVIP 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 276 FIPESPRW-LISGRFEAEVIRKAAKANGIVPSTIPDPSELQDLGSKKQOSH----- 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 FCPESPRFLINRNEENRAKSVLKLKRGTDV-----HDLOEMKEESROMMRE 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 330 ---NILDLRTWNIRMTTMSIMLMT-----ISVGYGLSLDTPNLHGDIFVNCFLSA 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 255 KVTILELFRSPAYROPILIAVVLQSLQOOLGINAVFYSTSI-----FEKA 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 MVEVPAY-----VLAWLLLOLYLPRR--YSMATALFLGSGVLLFMOLV---P 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 GVQOPVYATIGSIVNTAFTVVSFLFVVERAGRRTHLHGLAGMACQAILMTALALLQL 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 422 PDLYLATVLMVVGKFGVTAAS-----MVYYTAEIPTYVVRNMGVSVSTASRLGS-I 475
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 PMWSYLSIAI-----FGVAFVFGPGPIPFVAELESQGRPAIAVAGFSNNTSFI 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 476 LSPFYFYL---GAYDRFLPYILMGLSLTLTALTILF-LPESFGTPLPTIDQMLRVKM 530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 VGMCFQYVQLCGPY-----VFIIFVLLVLFIRYFKVPETKGRTFDEIASGFRGGAS 473
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 531 KIRKTP 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 QSDKTP 479
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-09-031-392-10
; Sequence 10, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031.392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkielejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-031-392-10

Query Match          7.0%  Score 203; DB 2; Length 493;
Best Local Similarity 25.0%  Pred No. 6.8e-13;
Matches 111; Conservative 76; Mismatches 147; Indels 110; Gaps 21;

QY 144 LTISLFFVGVLLGSFISQGLSDRGRK-----NVLFVTMGTMOTGFSFLOIFSKNFEMFV 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 LSVSIFAVGGMIGSFVGLXIGNLGRKXKAMLVNVLAIAGLLMG---LAKXAXSPFEMI 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 198 VLFVLGM-----GOISNYVAAFVLGTGLGKSVRIIFSTLGCVCIFYAFGYM 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 LGRFIIGLYCGLSSGVPMVMVGEISPTALRGALGT-----LNQLGIVIGILIA-Q 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 VLPFLFAYEIRD--WRMLLVALTMPGVLCVALMWFIPESPRWLI-----SQGRFEAEVIR 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 VLGLDSLILGNESLWPLLLGLTGVALLQLLLLPCEPSPRYLLINKNEERAKKALQRLR 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 299 KAAKANGIVPSTIPDPSELQD-----LSSKKQOSHNTLLRTWNIRMTTMSIMLMT- 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 GTADVQSEV-----AEMKDESXMXSEKXVS--VLELFRSRXYRQPVIIAIVLQLSQ 283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 354 -----ISVGYGLSLDTPNLHGDIFVNCFLSAMVEVPAY-----VLAWLLLO 395
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 QLSGINAVFYSTSI-----FEKAGVQPVYATIGAGVNVVFTVVSFVFE 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 396 YLPRYSMATALFLGG-----SVLLFMOLVPPDLYLATVLMVVGKFGVTAAS- - - - -MV 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 RAGRR--TLHLGLGGMACCAVLMTIALLDQVPMWSYVSIIVAFGEVAFVFGPGPIP 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 447 YVYTAELYPTVVRNMGVSVS-----TASRLGSLTSPYFV-YLGAYDRFLPYILMGLSITIL 501
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 389 WFIVAELFSQGRPAIAVAGFSNNTSNTFVGLLFQYIAELLGPY----VFIVFAVLLLL 444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 502 TAILT-LFLPESFGTPLPTIDQM 524
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 FFITFLKVPETKGR-----TFDEI 464
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-299-549-10
; Sequence 10, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,549

```

```

; FILING DATE: 26-APR-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-299-549-10

Query Match 7.0%; Score 203; DB 4; Length 493;
Best Local Similarity 25.0%; Pred. No. 6.8e-13;
Matches 111; Conservative 76; Mismatches 147; Indels 110; Gaps 21;

QY 144 LTISLFFVGLGSGISQSDRGK-----NVLFVTCMGQTGFSLQIFSKNFEMFV 197
DB 68 LSVSFAVGGMIGSLVXIGNIGRXXAMLVNVAIAAGLLMG---LAKXAXSEPMILI 124
QY 198 VLFVLGVM-----GQISNVAAFAVLGTGLGKSVRIIFSTGLGVCIFYAFGYM 244
DB 125 LGRFTIIGLYCGLSGVPMYGEISPTALRGALGT-----LNQLGIVIGILIA-Q 173
QY 245 VLPFLFAYFIRO--WRMLIVALTMPGLCVLAWMFIPESRWLI-----SQRFEAEVIR 298
DB 174 VLGLDSLGNESLWPLLLGLTGVPALLQLLLPCCPSRYLLINKNEEARAKKALQRLR 233
QY 299 KAANGANGIVPSTFDSELDQ----LSSKKQSHNLDLRTWNRMTYMSIMLWMT- 353
DB 234 GTADYSQEV-----AEMDESRRXXMXSEKXVS--VLELFRSRXXVQPVIIAIVLQLSQ 283
QY 354 -----ISVGYGLSLDTPNLHGDFVNCFLSAMVEVPAY-----VLAWLLLQ 395
DB 284 QLSGINAVFYVTSI-----FEKAGVGQPVYATIGAGVYVNTVTVSVFVVE 330
QY 396 YLPRYSMATALLFGG-----SVLLFMQLVPPDLYLATVLMVGKFGVTAAFS-----WV 446
DB 331 RAGRR--TLHLGLGGMAGCAVLMTIALALLDQVPMMSYVSIIVAFGVFAFFEVGPGPIP 388
QY 447 YVYTAELPTVVRNMGVGVSS-----TASRLGSLSPYFV-YLGAYDRFLPYILMGLTIL 501
DB 389 WFIVAELFSQGRPAIAIAGVGSNWTNFIIVGLLFQYIAELLGPY----VFIVFAVLLLL 444
QY 502 TAILT-LFLPESFGTLPDPTIDQM 524
DB 445 FIFTFLKVPETKGR----TFDEI 464

RESULT 14
US-08-647-397-4
; Sequence 4, Application US/08647397
; Patent No. 5972702
; GENERAL INFORMATION:
; APPLICANT: Beier, David R.
; APPLICANT: Brady, Kevin P.
; TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA

```

```

; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,397
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7048
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-647-397-4

Query Match 6.9%; Score 198; DB 2; Length 109;
Best Local Similarity 40.8%; Pred. No. 2.2e-13;
Matches 42; Conservative 22; Mismatches 39; Indels 0; Gaps 0;

QY 392 LLLOYLPRRYSMATALLFGGSLVLLFMQLVPPDLYLATVLMVGKFGVTAAFSMVYVYTA 451
DB 6 LSLVGLGHTTQAAALLAGGAILALTFPLDLQTVTVTLAVFGKGLSSFSCLFLYTS 65
QY 452 ELYPTVVRNMGVGVSSVTSASRLGSLSPYFVYLGAYDRFLPYIL 494
DB 66 ELYPTVIRQTGMGVSNLMTVRGSMVSLVKITGEVQFPFNII 108

RESULT 15
US-09-031-392-6
; Sequence 6, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283

```

REFERENCE/DOCKET NUMBER: 07334/072001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 509 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-031-392-6

Query Match 6.78; Score 193; DB 2; Length 509;
Best Local Similarity 24.18; Pred. No. 8.1e-12;
Matches 125; Conservative 93; Mismatches 175; Indels 126; Gaps 28;

QY 89 IANFSAL--GLEPGRVDL-----CQLQESCLDGV-----EFSQDYVLSITVENNLVC 136
Db 26 LAVFSAVIGLOFYNGIVINAPQVIEQSYNETWLGROGPEGSPPTGLTLW-----81
QY 137 EDDKRAPLITSLFFVGLSGFISQSLDRFGRK-----NVLFVTMGMTQGFSLQIFS 190
Db 82 -----ALSVAIFSVMISSEFLGIISQWLGKRAMLVNVLAVLGGSLMG---LANAA 132
QY 191 KNFEMFVFLVVG--MQQISNVYAAAFVLGTEILGKSVRIIFSTL-----GVCIFYAF 241
Db 133 ASYEMLIILGRELIGAYSLGSLVPMV--GETAPHLRGALGTNLQALVIGILIAQVL 190
QY 242 GYMVLPLFAYETRWMLLVALTPMGVLCVALMWFIPESPRWLTISQGRFEEAEVIRKAA 301
Db 191 GLESLGTASL--WPLLGLTVLPALLQLVLLPFCPEPRYLYI---IQNLGGPARKSL 244
QY 302 K-----ANGIVVPSTIFDPSELDLSSK--KQOSHNIIDLRTWNIRAVTMSINLWM 352
Db 245 KRLTGWADVSVL-----AELDKRKLDERPLSLQLLQSLRTHRQPLIAVVLQ 296
QY 353 T-----ISVGYFGLSLDTPNLHGDIENVCFLSAMVEVPAY-----VLAWLL 393
Db 297 SQQLSGINAVEYYSI-----PETAGVGPAYATIGAGVNVTVTLVSLL 343
QY 394 LQYLPFR--YSMATALFLGGSVLLFMQLV-----PPDLYLATVLMVVGKFGVTAAPS---444
Db 344 VERAGRRTLHLGLAGMCGCAILMTVALLLRLVPAMSVYSIVAI---FCGVAFPEIGP 399
QY 445 --MVVYVTAELYPTVVRNMGVGVSTASRLGS--ILSPYFVLGAYDRFLPYI-LMGSLLTI 500
Db 400 GPDPWFIVAELEFSQCPRAANAVAGFSNWTNFTIIGMGFQYVA--EAMGPYVLLFAVLL 457
QY 501 LPAITLFL--PESFGTLPDPTIDOMLRVGMKHKRTPS 537
Db 458 LGFFITFLRPETRGR---TFDQ---ISAAFRH-TPS 488

Search completed: August 16, 2001, 13:58:54
Job time: 216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:59:27 ; Search time 28.45 Seconds
(without alignments)
1491.361 Million cell updates/sec

Title: US-09-521-195-3
Perfect score: 2883
Sequence: 1 MRDYDEVTAFLGEGWPFQRL.....HTRMLKDGQRPRTILKSTAF 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 2883 | 100.0 | 557 | 2 JW0089 | organic cation tra |
| 2 | 2496 | 86.6 | 557 | 2 JE0346 | high-affinity carn |
| 3 | 777 | 27.0 | 576 | 2 T22509 | hypothetical prote |
| 4 | 733 | 25.4 | 593 | 2 JC4884 | organic cation tra |
| 5 | 712 | 24.7 | 556 | 2 S50862 | organic cation tra |
| 6 | 589 | 20.4 | 794 | 2 T27870 | hypothetical prote |
| 7 | 547 | 19.0 | 745 | 2 T16565 | hypothetical prote |
| 8 | 522 | 18.1 | 527 | 2 T01019 | transport protein |
| 9 | 483 | 16.8 | 539 | 2 C96758 | probable protein |
| 10 | 470.5 | 16.3 | 518 | 2 B86299 | hypothetical prote |
| 11 | 411 | 14.3 | 528 | 2 T21682 | hypothetical prote |
| 12 | 382.5 | 13.3 | 540 | 2 T25851 | hypothetical prote |
| 13 | 380.5 | 13.2 | 515 | 2 B96825 | hypothetical prote |
| 14 | 368.5 | 12.8 | 521 | 2 H86298 | hypothetical prote |
| 15 | 355.5 | 12.3 | 529 | 2 T23190 | hypothetical prote |
| 16 | 341.5 | 11.8 | 454 | 2 F75580 | probable sugar tra |
| 17 | 316 | 11.0 | 455 | 2 B83213 | probable MFS trans |
| 18 | 308 | 10.7 | 469 | 2 G65058 | hypothetical prote |
| 19 | 298.5 | 10.4 | 724 | 2 I50531 | transmembrane tran |
| 20 | 295.5 | 10.2 | 423 | 2 S74046 | probable sugar tra |
| 21 | 291.5 | 10.1 | 435 | 2 T15290 | hypothetical prote |
| 22 | 287 | 10.0 | 400 | 2 C69757 | transporter homolo |
| 23 | 287 | 10.0 | 422 | 2 G72234 | hypothetical prote |
| 24 | 279 | 9.7 | 524 | 2 T27082 | hypothetical prote |
| 25 | 274 | 9.5 | 683 | 2 S34961 | synaptic vesicle p |
| 26 | 272 | 9.4 | 478 | 2 T33985 | hypothetical prote |
| 27 | 271 | 9.4 | 480 | 2 T23608 | hypothetical prote |
| 28 | 267 | 9.3 | 443 | 2 E64725 | yaaU protein - Esc |
| 29 | 267 | 9.3 | 443 | 2 H85485 | probable transport |

ALIGNMENTS

RESULT 1

JW0089
organic cation transporter protein 2 - human
N:Alternate names: OCTN2
C:Species: Homo sapiens (man)
C:Date: 18-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 28-May-1999
C:Accession: JW0089
R:Wu, X.; Prasad, P.D.; Leibach, F.H.; Ganapathy, V.
Biochem. Biophys. Res. Commun. 246, 589-595, 1998
A:Title: cDNA sequence, transport function, and genomic organization of human OCTN2,
A:Reference number: JW0089; MUID:98289574
A:Accession: JW0089
A:Molecule type: mRNA
A:Residues: 1-557 <WUA>
A:Cross-references: GB:AF057164; NID:g3273740; PIDN:AAC24828.1; PID:g3273741
A:Experimental source: placenta
C:Comment: This transporter functions in the elimination of cationic drugs and other

Query Match 100.0%; Score 2883; DB 2; Length 557;
Best Local Similarity 100.0%; Pred. No. 7.8e-216;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| Qy | 1 | MRDYDEVTAFLGEGWPFQRLIFFLLSASTIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS | 60 |
| Db | 1 | DEVTAFLEGWGPFORLIFLLSASTIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS | 60 |
| Qy | 61 | AWRNHTVPLRLRDGVEVPHSCRRYRLATIANFSALGLEPGRDVLGLQLEQESCLDGWEFS | 120 |
| Db | 61 | WNTVPLRLRDGVEVPHSCRRYRLATIANFSALGLEPGRDVLGLQLEQESCLDGWEFS | 120 |
| Qy | 121 | QDVYLSTIVTEWNLVCEDDWKAPLTISLFFVGVLGSGFISGQSDRFGRKNVLFVTMGQ | 180 |
| Db | 121 | QVYLSTIVTEWNLVCEDDWKAPLTISLFFVGVLGSGFISGQSDRFGRKNVLFVTMGQ | 180 |
| Qy | 181 | TGFSFLQIFSKNFNFVFLVGVGQISNYVAARVLTGLKRSVRIIFSTLGVCIPIYA | 240 |
| Db | 181 | TGFSLQIFSKNFNFVFLVGVGQISNYVAARVLTGLKRSVRIIFSTLGVCIPIYA | 240 |
| Qy | 241 | FGYMWLPFAFYRIRDWRLMLVATMPGVLCVAMWFIPESPRWLISQGRFEAEVIRKA | 300 |
| Db | 241 | FGYMWLPFAFYRIRDWRLMLVATMPGVLCVAMWFIPESPRWLISQGRFEAEVIRKA | 300 |
| Qy | 301 | AKANGIVVPSTIFDPSQLSSKKQKQSHNILDLLRTNIRMTIMSLMNTISVGVFG | 360 |
| Db | 301 | AKANGIVVPSTIFDPSQLSSKKQKQSHNILDLLRTNIRMTIMSLMNTISVGVFG | 360 |
| Qy | 361 | LSLDFPNLHGDIFVNCFLISAMVEVPAYVLAWLLQYLPRRYSMATALFLGGSVLLFMQLV | 420 |
| Db | 361 | LSLDFPNLHGDIFVNCFLISAMVEVPAYVLAWLLQYLPRRYSMATALFLGGSVLLFMQLV | 420 |
| Qy | 421 | PPDLVYLATVLMVGKFGVTAAFSMVYVVTABLYPTVVRNMGVSGVSTASRLGSLSPYF | 480 |

hypothetical prote
hypothetical prote
probable transport
probable MFS trans
synaptic vesicle p
metabolite transpo
hypothetical metab
hypothetical prote
arabinose-proton s
hypothetical prote
sugar transporter
probable transport
metabolite transpo
hypothetical prote
metabolite transpo

Db 421 PPDLVYLATVLMVGVKFGVTAFAEMVYVTAELPTVVRNMGVSVSTASRLGSLSPYF 480
QY 481 VYLGAYDRFLPYLIMGSLTILTAITLFLPESFGTLPDIDOMLRVKMKHKRTPSHTR 540
Db 481 VYLGAYDRFLPYLIMGSLTILTAITLFLPESFGTLPDIDOMLRVKMKHKRTPSHTR 540
QY 541 MLKDGQERTILKSTAF 557
Db 541 MLKDGQERTILKSTAF 557

RESULT 2
JE0346
high-affinity carnitine transporter, CTL - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JE0346
R: Sekine, T.; Kusuhabara, H.; Utsunomiya-Tate, N.; Tsuda, M.; Sugiyama, Y.; Kanai, Y.; Endo
Biochem. Biophys. Res. Commun. 251, 586-591, 1998
A:Title: Molecular cloning and characterization of high-affinity carnitine transporter f
A:Reference number: JE0346; MUID:99011422
A:Accession: JE0346
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-557 <SEQ>
A:Cross-references: DDBJ:AB017260; NID:g3869208; PIDN:BA034399.1; PID:g3869209

Query Match 86.6%; Score 2496; DB 2; Length 557;
Best Local Similarity 85.3%; Pred. No. 7.8e-186;
Matches 475; Conservative 42; Mismatches 40; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEMGPFQRLIFLLSASIIPIGFTGLSSVFLIATPEHRCRVPDAAANLSS 60
Db 1 MRDYDEVTAFLGEMGPFQRLIFLLSASIIPIGFTGLSSVFLIATPEHRCRVPDAAANLSS 60
QY 61 AWENHVPRLRDRGVRPHSCRRYRLATIANFSALEGRDVLGLOEQESCLDGEWFS 120
Db 61 AMNHSIPLETKDGRVPOSCRRYRLATIANFSALEGRDVLGLOEQESCLDGEWFS 120
QY 121 QVYLSTIVTEWNLVCEDDWKAPLITISFFVGVLLGSFISGQLSDRGKKNVLFVTMGQ 180
Db 121 KOVFLSTIVTEWNLVCEDDWKAPLITISFFVGVLLGSFISGQLSDRGKKNVLFVTMGQ 180
QY 181 TGFSLQIFSKNFEMFVVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIIFYA 240
Db 181 TGFSLQIFSKNFEMFVVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIIFYA 240
QY 241 FGVMLPLPAYFIRDRMRLVLTWPGVLCVLMWFIPEPRLISQGRFEAEVIRKAA 300
Db 241 FGVMLPLPAYFIRDRMRLVLTWPGVLCVLMWFIPEPRLISQGRFEAEVIRKAA 300
QY 301 AKANGIVPSTIFDPSELQDLSKKQSHNIDLLRTNIRMTIMSLMWTISVGYFG 360
Db 301 AKANGIVPSTIFDPSELQDLSKKQSHNIDLLRTNIRMTIMSLMWTISVGYFG 360
QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLF 420
Db 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLF 420
QY 421 PPDLVYLATVLMVGVKFGVTAFAEMVYVTAELPTVVRNMGVSVSTASRLGSLSPYF 480
Db 421 PSELVYSLTALVMVGVKFGVTAFAEMVYVTAELPTVVRNMGVSVSTASRLGSLSPYF 480
QY 481 VYLGAYDRFLPYLIMGSLTILTAITLFLPESFGTLPDIDOMLRVKMKHKRTPSHTR 540
Db 481 VYLGAYDRFLPYLIMGSLTILTAITLFLPESFGTLPDIDOMLRVKMKHKRTPSHTR 540
QY 541 MLKDGQERTILKSTAF 557
Db 541 TOKDGESPTVLKSTAF 557

RESULT 3
T22509
hypotheical protein F52F12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22509
R: Matthews, L.
submitted to the EMBL Data Library, December 1996
A:Reference number: Z19573
A:Accession: T22509
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-576 <WIL>
A:Cross-references: EMBL:Z83228; PIDN:CA05732.1; GSPDB:GN00019; CESP:F52F12.1
A:Experimental source: Clone F52F12
C:Genetics:
A:Gene: CESP:F52F12.1
A:Map position: 1
A:Introns: 16/1; 23/3; 50/1; 80/3; 108/2; 134/3; 221/1; 257/2; 319/3; 358/2; 490/3; 5

Query Match 27.0%; Score 777; DB 2; Length 576;
Best Local Similarity 31.4%; Pred. No. 1.5e-52;
Matches 165; Conservative 118; Mismatches 225; Indels 18; Gaps 6;

QY 3 DYDEVTAFLGEMGPFQRLIFLLS-ASIIPIGFTGLSSVFLIATPEHRCRVPDAAANLSSA 61
Db 16 DFDFVLEQVGNVGYQIVFFFIICLPTLSAFSAFNIPFVGNPPHTCHIPGKEYLRP 75
QY 62 WRNHTVPLRLDRGVRPHSCRRYRLATIANFSALEGRDVLGLOEQESCLDGEWFSQ 121
Db 76 LTNDTQIL-----SCKQYNETQINVFRAFTSAP-VDTYSDRISLVPCCQNGWDYN 124
QY 122 DVYLSTIVTEWNLVCEDDWKAPLITISFFVGVLLGSFISGQLSDRGKKNVLFVTMGQ 181
Db 125 STYLDLSLTFEFLNVCDDQAWIEISFFYVGSFIGNCLFGYVADKFGRRSRFFVILT 184
QY 182 GFSFLOIFSKNFEMFVVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIIFYA 241
Db 185 VCGTASSFADIESFIILRFFETGLAPFALFQIFPICMEFPMGNSGR-IFSLGTLSTFFGA 243
QY 242 GYVMLPLPAYFIRDRMRLVLTWPGVLCVLMWFIPEPRLISQGRFEAEVIRKAA 301
Db 244 AMALLGVVAFIRRMQLTFFCNAPFAFYIYFFELPESRPSWSVSGKNADAKKLLK 303
QY 302 KANG---IVVPSTIFDPSELQDLSKKQ--OSHNIIDLLRTNIRMTIMSLMWTISV 356
Db 304 KMNGKSNVDVDELVDMSKMHQNAEKEETKRSNVTDLFTPNLRKTLIVTIWVNAI 363
QY 357 GYFGLSDTPNLHGDIFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLF 416
Db 364 IYNGLTNLVSNLPPDDYWSFIINGAVELPGYFVVMPLLCACAGRRWTLAATMIVCGIGCVS 423
QY 417 MQLVPPDLYLATVLMVGVKFGVTAFAEMVYVTAELPTVVRNMGVSVSTASRLGSL 476
Db 424 AMFPDGYPLVLSASFTIGKFGVGSFVAVIIFAGELYPTVVRNMGVSVSTASRLGSL 483
QY 477 SPFVYLGAYDRFLPYLIMGSLTILTAITLFLPESFGTLPDID 522
Db 484 APHTVNLGKIVKILPLLIMGLMALSAGILFTFLPETIGAPLPMTIE 529

RESULT 4
JC4884
organic cation transporter protein 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
C:Accession: JC4884
R: Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.
Biochem. Biophys. Res. Commun. 224, 500-507, 1996
A:Title: cDNA cloning and functional expression of a novel rat kidney organic cation
A:Reference number: JC4884; MUID:96295517

A:Accession: JC4884

A:Molecule type: mRNA

A:Residues: 1-593 <OKU>

A:Cross-references: DBJ:D83044; NID:gl502282; PIDN:BAAL1754.1; PID:d1012421; PID:gl5022

A:Experimental source: kidney

C:Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 25.4%; Score 733; DB 2; Length 593;
Best Local Similarity 35.2%; Pred. No. 3.9e-49;
Matches 198; Conservative 89; Mismatches 234; Indels 42; Gaps 15;

QY 1 MRDYDEVTAFLGSGWGFQRLIFLLSASITPNGTGL--SSVFLATPEHRCRVPDAANL 58

Db 1 MSTVDDILEHGFHFQKQTFELL--ALLSCAFPIYGVIFLGFTHDHCWSGAACL 58

QY 59 SS--AWR-----NHTVP-LRLRDGVRPHSCRRYR-----LATANFSAFLGLEPGRDV 103

Db 59 SQRCSQAELNYTPVGLGPSDEASFLSQCMRYEVDWNOQSTLDCVPLSLA-----A 112

QY 104 DLQLEQESCLDGEFSQDVYLSITVTENLVCEDDWKAPLISLFFVGVLLGSFISQGL 163

Db 113 DRNQLPLGCEHWGYYNTEG--SSIVTEFNLCVCAHSMLLDFQSVYVNGFFIGAMMIGYL 170

QY 164 SDRFGKRVLFVTMGMTGFSFLQIFSKNFEMFVVLFLVGMGQISNYVAAPVLGTEILG 223

Db 171 ADRFGKRVLFVTMGMTGFSFLQIFSKNFEMFVVLFLVGMGQISNYVAAPVLGTEILG 230

QY 224 KSVRIIFSTLGVGCIYAF--GYMVLPLFAFYIRDRWMLLVALTMPLGVALWFWFIPESP 281

Db 231 LGYR--RWGICYAFTVGLLILAGVAVIPNWRWQLQFVTLNFCFLYFWCIPESP 287

QY 282 RWLSQGRPEAEVIRKAANGANGIVVPSITFDPSLQSLSSKKQOOSHNLILLRTWNR 341

Db 288 RWLSQKIVKAKIITKIAKKNGKSPVSLQNLTPDEDAKKLKP--ILDLVTPQR 345

QY 342 MVTIMSLMWTISVGYFGLSLDTPNLHGD-IFVNCFLSAMVEVPAYVLAWLLOLPLRR 400

Db 346 KHTLIMYNFTSSVLYQGLIMHM-GLAGNIYLDFFYSALVEFPAAFIITIDRVGR 404

QY 401 YSMATLFGGSLVLLFMQLPDPXYLATVLMVGMGFGVTAFAFSVYVYTAELVTVSRN 460

Db 405 YPWAWSMWAGAACLASVFPDDQLWKITIACLGRMGITMAYEMVCLVNAELPYIRN 464

QY 461 MGCVSSTASRLSGISLSPFYV-LGAYDRFLPYLMGSLTILTAITLPLPSFGTPLPD 519

Db 465 LGVLVCSMCDIGGIITPFLVRLTDIWMFEPLVFAVGLVAGALVLLPETKALPE 524

QY 520 TID-----QMLRVKGMKHKRTPS 537

Db 525 TIEDAENQPRKRENLPSPS 547

RESULT 5
S50862
organic cation transport protein Ocri - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 05-Nov-1999

C:Accession: S50862; S78533; I58089

R:Gruendemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.

Nature 372, 549-552, 1994

A:Title: Drug excretion mediated by a new prototype of polyspecific transporter.

A:Reference number: I58089; MUID:95082907

A:Accession: S50862

A:Molecule type: mRNA

A:Residues: 1-556 <GRU>

A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622

A:Experimental source: kidney

R:Gorboulev, V.G.

submitted to the EMBL Data Library, January 1995

A:Reference number: S78533

A:Accession: S78533

A:Molecule type: mRNA

A:Residues: 1-342, 'N', 344-556 <GOR>

A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622

C:Keywords: glycoprotein; phosphoprotein; transmembrane protein

F:20-46/Domain: transmembrane #status predicted <TM1>

F:154-171/Domain: transmembrane #status predicted <TM2>

F:178-197/Domain: transmembrane #status predicted <TM3>

F:243-260/Domain: transmembrane #status predicted <TM4>

F:267-283/Domain: transmembrane #status predicted <TM5>

F:350-366/Domain: transmembrane #status predicted <TM6>

F:380-398/Domain: transmembrane #status predicted <TM7>

F:406-425/Domain: transmembrane #status predicted <TM9>

F:435-452/Domain: transmembrane #status predicted <TM10>

F:469-485/Domain: transmembrane #status predicted <TM11>

F:494-514/Domain: transmembrane #status predicted <TM12>

F:71, 97, 113, 432/Binding site: carboxylate (Asp) (covalent) #status predicted

F:286, 292/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pred

F:296, 343, 550/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status

Query Match 24.7%; Score 712; DB 2; Length 556;
Best Local Similarity 33.8%; Pred. No. 1.5e-47;
Matches 193; Conservative 96; Mismatches 230; Indels 52; Gaps 14;

QY 1 MRDYDEVTAFLGSGWGFQRLIFLL--SASITPNGTGLSSVFLATPEHRCRVPDAAN 57

Db 1 MPTVDDVLEQVGEFGWFOKQAFLLCLLISASLAPI-YVGI--VFLGTPGHYCNQPGVAE 57

QY 58 LSS--AWR-----NHTVP-LRLRDGVRPHSCRRYL-----ATTANFSA 95

Db 58 LSQRCSQAELNYTPVGLGPSDEASFLSQCMRYEVDWNOQSTLDCVPLSLVANSRQL 117

QY 96 GLEFGRDVLDLQLEQESCLDGEFSQDVYLSITVTENLVCEDDWKAPLITSLFVGVLL 155

Db 118 PLGP-----CEHWY--DTPGSSIVTEFNLCVGDWAKVDLFQSCVNLGFFL 162

QY 156 GSFISGQLSDRFGKRVLFVTMGMTGFSFLQIFSKNFEMFVVLFLVGMGQISNYVAAF 215

Db 163 GSVLVGYIADFRGKRLCLLVTLVTSVSGVLTAVADYTSMLLFRLLQGMVSKGWSVGY 222

QY 216 VLGTEILGKSVRIIFSTLGVGCIYAF--GYMVLPLFAFYIRDRWMLLVALTMPLGVAL 273

Db 223 TLITFVSGSYR--RTTAILYQMAFTVGLVGLAGVAYAPIDRWMLQLAVSLFTFLFLY 279

QY 274 WVFIPESRWLLISQGRPEAEVIRKAANGANGIVVPSITFDPSLQSLSSKKQOOSHNL 333

Db 280 YWVFPESRWLLISQGRPEAEVIRKAANGANGIVVPSITFDPSLQSLSSKKQOOSHNL 337

QY 334 LLRTWNIIRMTIMSLMWTISVGYFGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWL 393

Db 338 LFRTPTLRKHVILMYLWFSQAVLYQGLIMHVGATGANLYLDDFFYSILVEPFAAFILVT 397

QY 394 LOYLPRRYSMATALFLGGSVLLFMQLPDPXYLATVLMVGMGFGVTAFAFSVYVYTAEL 453

Db 398 IDRIGRIYPIAASNLVITGAACLLMIFIPHELHNLVTLACLRMGATITVLQMVCLVNAEL 457

QY 454 YPTVVRNMGVSVSTASRLSGISLSPFYV-LGAYDRFLPYLMGSLTILTAITLPLPES 512

Db 458 YPTFTRNLGMVCSALDGLGIFTFPFWRLMEVWQALPLILFGVLGTAGATILLPET 517

QY 513 FGCTPLPTIDQMLRVKGMKHKRTPSHRMLK 543

Db 518 KGVALPETIEEAENL-GRKSKAKENTIVLQ 547

RESULT 6

T27870

hypothetical protein ZK455.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T27870

R:White, S.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z20432

A:Accession: T27870
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-794 <WIL>
 A:Cross-references: EMBL:Z66567; PIDN:CAA91492.2; GSPDB:GN00028; CESP:ZK455.8
 A:Experimental source: clone ZK455
 C:Genetics:
 A:Gene: CESP:ZK455.8
 A:Map position: X
 A:Introns: 52/3; 119/2; 158/1; 207/3; 291/3; 335/2; 428/3; 610/2; 636/3; 733/3

Query Match 20.4%; Score 589; DB 2; Length 794;
 Best Local Similarity 27.8%; Pred. No. 8e-38;
 Matches 167; Conservative 113; Mismatches 231; Indels 90; Gaps 22;
 QY 3 DYDE---VTAFLGEMPPFORLIFLLSASIIIPNGFTGLSSVFLIATPEHRCVPE---D 54
 Db 155 DFEEDDLGILQGGCYWQIWIYLIISVQOVPHAMENUSVVTMYTQPHWCKIPFFNEE 214
 QY 55 AANLSSAWRNHT-----VPLRLDRGREVPH--SCRRYLATIANFSALEPGGRD 102
 Db 215 SFSAEGLTYNTWDVNLSTIAFPRTFNKORNELHHDQCHYPE---RDVYHKLSPWAQ 270
 QY 103 V-DLGQLEQESCLDGFESQDYLSTIVTENLVCEDDKAPLITISLFFVGLLGSFISG 161
 Db 271 VKDNATGKINRCKEWDYTSVMDRTIVTWNRVCDNNMSRAHVHMSYSLGLVGCFCVG 330
 QY 162 QLSDRFRGRKNVLEVTMGMTG--FESLQIFSKNPEMF--VVLVFLVGMGTQSNVAAEVLGT 219
 Db 331 FISDRYGRKTAI--TGFGILLTMLFGELLTYSKEFEILVVRFLAATNEAAD--LAAYVLCM 388
 QY 220 EILGKSVRIITFSLGVCIFYAFGYVWLPPLFAFPIRDMRLVALTMGVLCAVLMWFIPE 279
 Db 389 EVTGTRYRSIVGSL--IQAPWACGYAFALIAIYLTAKSWTHILICVLLHIIISMLLYFLPE 447
 QY 280 SPWLLISQGFEEAEVILIRKAAKANGIVVPSTIFDSELDLSSKKQKQSHN---ILDLL 335
 Db 448 SPWLLLNKTKQREKILIRACHYNSRLPS---DLGLVRHAEEKKMKHMKPEKPSYFHLF 504
 QY 336 RTWNRMTVIMSLMW-----TISVGY-----FGLSLDT 365
 Db 505 RSSELFRNVLVFWIQKNIFONSRFQSDNSCGLLRNGYCVIRSVITRKVGRTLSE 564
 QY 366 P-----NLDGIFVNCFLSAMVEPAYVLANLLQYLPRYSMATALFLGGSVL 414
 Db 565 KFYRNOKLCFRVRFDGNFTLNAMAGAEILTLVFCVFLLR--MGRKRSQMLVLF--GSGILF 622
 QY 415 LFMQLVPPDLY-----LATVLVMGKGVTAFAFMSVYVYTAELYPTVVRNMGVSSSTA 469
 Db 623 LLTSVV---WVYRKQSTLALIFMLLSKACIQGSFNILYIFTSELNPTVVRNMGVSSSTA 679
 QY 470 SRIGSTLSPYFVYLGAYDR--FLPYILMGSLLTILTAITLFLPESFGTLPDPTI--DQMLRV 527
 Db 680 ARMGAGASGYIALSDVTPLVPMVMTIFACFSLLAGCLVLLPETOGLPLPDTILDSVQMV 739
 QY 528 K 528
 Db 740 K 740

RESULT 7
 T16565
 hypothetical protein K05F1.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16565
 R:Wohldmann, P.
 submitted to the EMBL Data Library, June 1995
 A:Description: The sequence of C. elegans cosmid K05F1.
 A:Reference number: Z18537
 A:Accession: T16565
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-745 <WOH>
 A:Cross-references: EMBL:U29377; NID:9868173; PID:9868176; PIDN:AAA68713.1; CESP:K05F
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:K05F1.6
 A:Introns: 70/2; 160/3; 205/3; 249/3; 293/2; 350/1; 386/3; 462/3; 546/2; 669/3; 680/3

Query Match 19.0%; Score 547.5; DB 2; Length 745;
 Best Local Similarity 27.5%; Pred. No. 1.2e-34;
 Matches 165; Conservative 105; Mismatches 243; Indels 87; Gaps 17;
 QY 3 DYDEVTAFLGEMPPFORLIFLLSASIIIPNGFTGLSSVFLIATPEHRCVPE--- 53
 Db 112 DFEGLNLIIGCRWQIWIYLIQLOIPHAMFNUNVMMYDPEFCWCPGFNDTNT 171
 QY 54 -----DAANLSSAWRNHTVPLRLDRGREVPHSC-----RRYLATIANFSA 94
 Db 172 ISSHGPYLMGVDDIKNISVFEPNAN-----SDGAYQRDCVYFSEERYQLRRMPLT 226
 QY 95 LGLEPGRDVLGQLEQESCLDGFESQDYLSTIVTENLVCEDDKAPLITISLFF--VGV 153
 Db 227 AMSEAWKDV---APKKKC-QAYHFEKDVMTETIVTDENLVC--DSWFAKGAHMFYSIGY 280
 QY 154 ILGSPISQGLSDRFRGRKNVLEVTMGMTGFSFLQI-----FSKNFEMFVVLVFLVGMG 206
 Db 281 LLGCVLGGIADSKIGRKPTII-----GFGILSSMLGVLPFPNDYPMFELLIRLSAIC 333
 QY 207 QISNVAAFLVTEILGKSVRIITFSLGVCIFYAFGYVWLPPLFAFPIRDMRLVALTM 266
 Db 334 NEAADLAAITTCMEITGTGYRAMVGS--LQAPWALGYALLAIYLTAKSWTKIQTIVAA 392
 QY 267 GVLCAVLMWFIPESPRWLISQGFEEAEVILIRKAAKANGIVVPSTIFD-----PS 316
 Db 393 HFSIIFTCSPESPRWLNQNRVSEAEVIRKACRE-----PPFPNMCTSKGNLPS 447
 QY 317 ELQDLSKKKQSH-----NILDLRTNIRMTVIMSLMWITISVYFGLSL-----D 364
 Db 448 DLELVSHREKLNKKGKIGFLDFTMKELRYRTISVCIVFNATLVYGLVWALSQDS 507
 QY 365 TPN---LHGDIFVNCFLSAMVEPAYVLANLLQYLPRYSMATALFLGGSVLLFMQLVP 421
 Db 508 AGRTLTGYTHLNLNGTAGAIEPTLFACVMMQOLGRKALMLTLITSLGLFIIVAMLSMV 567
 QY 422 PDLYLATVLVMGKGVTAFAFMSVYVYTAELYPTVVRNMGVSSSTAFLGSLTSLPYEV 481
 Db 568 SCHYMLALAFMYFKIAVQGFANILYIFTSELYPTVVRNMGVSSSTAFLGSLTSLPYEV 627
 QY 482 YLGAYD--RFLPYILMGSLLTILTAITLFLPESFGTLPDPTIDQMLR--VKGMRKRTPSHT 539
 Db 628 LLSNISLPIVPMIIFAVFSLFAGMLVFVPETSEKLPETLDDAINFLEPTKQFKTIEST 687

RESULT 8
 T01019
 transport protein homolog YUP8H12R.2 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 22-Oct-1999
 C:Accession: T01019
 R:Theologis, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K
 Oefner, P.; Davis, R.W.
 submitted to the EMBL Data Library, May 1998
 A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.
 A:Reference number: Z14227
 A:Accession: T01019
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-527 <THE>
 A:Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152576; GSPDB:GN00059; ATSP:YU
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Gene: ATSP:YUP8H12R.2

RESULT 9
C96758
probable protein transporter Tl8K17.11 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96758
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso-
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.R.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96758
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-539 <STO>
A:Cross-references: GB:AE005173; NID:g6598860; PIDN:AAF18714.1; GSPDB:GN00141
C:Genetics:
A:Gene: Tl8K17.11
A:Map position: 1

RESULT 10

B86299 hypothetical protein AAD34691.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86299
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.P.; Hughes, B.; Huizlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marfia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B86299
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-518 <STO>
A:Cross-references: GB:AE0051172; NID:94966360; PIDN:AAD34691.1; GSPDB:CN00141
C:Genetics:
A:Map position: 1

Query Match 16.3%; Score 470.5; DB 2; Length 518;
Best Local Similarity 28.2%; Pred. No. 7.4e-29; Mismatches 218; Indels 35; Gaps 12;
Matches 136; Conservative 94; #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

QY 55 AANLSSAW---RNHTVPLRLRDGREVPHSCRRYRLATIANFSAALGLEGPRDVLGQLEOE 111
DB 44 AALVSFAFEDQAQTFITVFDTSQPMWHCONSDRVQSVCTSSNL-----89

QY 112 SCL---DGEWESQDYLSTIVTENLVCEDDWKAPLISLFFVGVLLGSFISQSLSD-RF 167
DB 90 -CTLPNOTWSMDLPHVS-IIEKGLQACAGSLKGFPPASSFFLGLGGLTGLADSS 147

QY 168 GRKNVLFVTMGOTGFSFLOFSKNFEMFVFLVFLVGMGOISNVAAVFLGTILGKSVR 227
DB 148 GRKNMLLSCLIMSLSMLTAFSTSIWYAFRLFLNCGGRATGTCTALVSTELVKKWR 207

QY 228 IFSTLGVCIYFAF---GYWVLPPLFAYFI--RDWMLLVALTMPGVLCVVALWVFIPESR 282
DB 208 --GQVGAMGFCFTLGLFSLPMLGYINEGNSRNLVYVTSIPTLIYCCLVRSFVRESR 264

QY 283 WLISOGREFEEAVIRKAAKANGIVVPTIFDPSELQDLSKKQKQSHNILDLL--RTWNI 340
DB 265 WLIVKGRKEEAVSLQSTA-SNAITMSEFNLCEFEVNDQSKNPVDYDALKILVRKSWF 323

QY 341 RMYTMSIMLWMTISVGVGLSLDTPNLHGDIFVNCFLSAMVEYPAVYLAWLLLOLPRR 400
DB 324 RRL-LAAMVVGFGIGVYGYMPLALTNLNPLVGVVFNALSEFPAPLITFFFDIKINR 382

QY 401 YSMATALLFGGSVLLFMQLVPPDLVYLVGMKGVTAAPSMVYVTAELYPVVRN 460
DB 383 DALIGFTALSGLSALIAVGLQQLGSLQVLELVSVFSACTAFNMTLIYTIEMFPTCVRN 442

QY 461 MGVGVSSTASRLGTLSPFYVILGAYDRFLPYILMGLSLITAILTLFLPESFGTLPDT 520
DB 443 SAISMVQALVFGVFSVPMVAAQRENGFWSYGLFGLIGLGLFVGLPETRGSVLCDT 502

QY 521 IDQ 523
DB 503 MDE 505

RESULT 11
T21682
hypothetical protein F32H5.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21682
R:Steward, C.
submitted to the EMBL Data Library, November 1996
A:Reference number: 219458
A:Accession: T21682
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-528 <WIL>
A:Cross-references: EMBL:281524; PIDN:CAB04250.1; GSPDB:GN00023; CESP:F32H5.4
A:Experimental source: clone F32H5
C:Genetics:
A:Gene: CESP:F32H5.4
A:Map position: 5
A:Introns: 63/1; 108/3; 148/3; 249/2; 408/1; 472/3

Query Match 14.3%; Score 411; DB 2; Length 528;
Best Local Similarity 26.4%; Pred. No. 3.1e-24;
Matches 126; Conservative 91; Mismatches 169; Indels 92; Gaps 18;

QY 100 GRDVLGQLEQESCLDGEF-----SQDYLSLTIYVENLVCEDDWKAPLISLF 149
DB 73 GPDVFI---KDICKSSDFDHPNPNCTITSKYDYF--SINVDYGHFCGEGAWKTSISVQ 126

QY 150 FVGVLLGSFISQSLSDRFGRKNVLFVTMGOTGFSFLOFSKNFEMFVFLVFLVGM--G 206
DB 127 MVGVLLGSVTSVAVDRYGRUKLVSVCFEMVSSLSILNTFAKDLIVFTIIRTLSIFKGG 186

QY 207 QISNVAAVFLVGTILGKSVRLIFSTLGVCIYFAVGMVLPPLPAYFIRDW---RMLLVAL 263
DB 187 LLSTY---GVTKMEHVPQROHFWIATM---ISWAPMILSFVAYLCHDWTYQYAFAL 240

QY 264 TMPGVLCVALWVFIPESPRWLISOGREFEEAEVIRKAAKANGIV-----VPSTIFDPS 316
DB 241 SNPQAV---VPLFKESPRMLIQAGRIEAEARQVLKHIMEVDGNTSDHSNNEIEMLOTEK 297

QY 317 ELQDLSKKQKQSHNILDRLTNIRMYTMSIMLM---TISVGVFGSLDTPNLHGDIF 373
DB 298 KRQERIKRKNYDFRHLF--WNKYMASV-TMILWLMFSTFTNYGFFVNEIKLSGLY 354

QY 374 VNCFLSAMVEYPAVLAWLLLOLQYLPYRSMATALLFGGSVLLFMQLVPPDLVYLAWLV 433
DB 355 INALMGS-----LRWL-----NIFGLADLKAFLGRKHILHLSKLTIT 395

QY 434 VKRFGV-----TAAFSMVYVYTA-----ELYPTVVRNMGVGS 466
DB 396 ICVFSIFVTVYFEYEDYSLIIRVATLLASATASQVFIKSMVLMFEYPTVIRNSAVSEK 455

QY 467 STASRLGSILSPYVILGAYDRFLPYILMGLSLITLAI-LTFLPESFGTLPDITDQ 523
DB 456 SSASRIGTILGPQLFICPY-KSLPYAILTGFCFLDAIAFQALRPETKGRPLPETMPE 512

RESULT 12
T25851
hypothetical protein T01B11.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25851
R:Geisler, C.; Stellyes, L.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid T01B11.
A:Reference number: Z20099
A:Accession: T25851
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-540 <GEI>
A:Cross-references: EMBL:U80931; PIDN:AAB38002.1; GSPDB:GN00022; CESP:T01B11.5
A:Experimental source: strain Bristol N2; clone T01B11
C:Genetics:
A:Gene: CESP:T01B11.5
A:Map position: 4
A:Introns: 85/1; 128/3; 209/2; 269/2; 352/2; 381/3; 489/3

Query Match 13.3%; Score 382.5; DB 2; Length 540;
Best Local Similarity 26.6%; Pred. No. 5.2e-22;
Matches 120; Conservative 78; Mismatches 178; Indels 75; Gaps 14;

QY 127 TIVTENLVCEDDWKAPLITSLFFVGVLLGSFISQSLSDRFGRKNVLFVTMGOTGFSFL 186
DB 124 SVQVEFNVICDARKVKNTITVQTFGLVGAATFGQVSDTFGRKALLISTLGLNGLFNWI 183

QY 187 QIFSKNFEMFVFLVFLVGMGOISNVAAVFLGTILGKSVRIIFSSTLGVCIYFAVGMVL 246
DB 184 TAYSPDLFYMVWRTLAGVFTGGVTVVQVFMVENIPRKRDMIQN---SIWSPNLIIF 240

QY 247 PLFAYFIRDWRMMLLVALTMPGVLCVVALWMTIPESPRWLISQGRFEAE---VIRKAAKA 303
DB 241 PFVAWLQCDWRTMSVWIAASSIATFLACLILESPRWLIQKGRIEEARKSLIRKTRD- 299

QY 304 NGIVVPSPIF-----PSELQDLSKKQKQSHNILDRLTNWNI--RMVTMSI 348
DB 300 -----LYDETEPEKOLDEVLHVEAEKHARSSKKAKKYTFIHLCTWKMIAQSFTFG 351

QY 349 MLWMTISVGVFGSLDTPNLHGDIFVNC-----FLSAMVEYPAVYLAWLLLOLPRR 400
DB 352 ITCTTFTIV--YSLMYNNERKLSGLYWNLAIMGASRIINIAVSADIYKLPWFGRKMI-NQ 408

QY 401 YSMATALLFGGSVLLFMQLVPPDLV---YLATVLMVGMKGVTAAPSMV-----YVYTA 451

Db 409 IAMICTLFA-----LIMIAVLYFYNGQLIAI-----GTVAASVCMCSOLFIAKYMVN 457
QY 452 ELYPTVVRNMGVSVSTASRLGSLTSPFYVYLGAIDRFLPYILMGLSLTILT-AILTFLFLP 510
Db 458 ELYPTAVRNLAASVATWSRIGKSFQSLFYLSDYAEWIPYAVLFSQLYDFIVLSLFLP 517
QY 511 ESFGTPLPDTIDQMLRVKMGKHKRTTPSHTRM 541
Db 518 ETKGVL-----ENHLPPKHRI 535
RESULT 13
B96825
hypothetical protein T8K14.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B96825
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B96825
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-515 <STO>
A:Cross-references: GB:AE005173; NID:g4835768; PIDN:AAD30235.1; GSPDB:GN00141
A:Gene: T8K14.17
A:Map position: 1
Query Match 13.2%; Score 380.5; DB 2; Length 515;
Best Local Similarity 26.7%; Pred. No. 7e-22;
Matches 118; Conservative 95; Mismatches 186; Indels 43; Gaps 15;
QY 113 CLD-----GHEFSQDYVLTITVENLVCEDDWKAPLTISLFEVGVLL 155
Db 72 CLDHTICNPATTDICKIPRSAMDGGFGKSVISEFDECCSSFLRSLPSSSTFYVGSIV 131
QY 156 GSFISQSLD-RFGKKNVLT--MGMTGFSFLQIFSKNFEMFVFLVGVGMQISNY 211
Db 132 GGVLAMIPDGLGRKQLLFFSSFMSL-TGISI--FLSSNIWIYSFLKFVIGFARSQTG 188
QY 212 VAAVVLGTEILGKSVRIITSLGVCFYAFGVYMLPLFAYFIR--DWRMLVALTMP-GV 268
Db 189 TVALVLISERISTKWR-PRATVPTFLFVGLMSLGLIAYLRHASRWKVLCTSIAPGI 247
QY 269 LCVALWVFPSPRWLISOGREAEVIRKAAKANGIVPSTIFDPSELQDLSKKQ-Q 327
Db 248 HSIFYFFALESPRHLHLEGNKEAEVLKRISPNRNGVLESV----SRLRPKEFLEQTS 304
QY 328 SHNIDL-LRTNIRMTTMSLWMTISVGYGLSLDTPNLHGDFVNCFLSAMVEVPA 386
Db 305 SVSKDLFIKNAFRVILVIMF-GLGMSYGYVPLAVRDIKVNIYKSEALNANVELPT 363
QY 387 YVLAWLLQYLPRLRYSMATFLGGS--VLLFMQLVPPDLY---YLATVLMVGVKFGVTA 441
Db 364 FVVTPILEQFSRRSVLVNCLIGGASGVLCFV-----SLYGRTRIAFALELGSPFCARI 419
QY 442 AFSWVYVYTAELPTVVRNMGVSVSTASRLGSLTSPFYVYLGAIDRFLPYILMGLSLIL 501
Db 420 GFNLAIYVELFPTCVRSATMMLRQALVVGACCPLIASLGSRNVPSLSFVAFGAMSG 479
QY 502 TAILTLFPESTGPLPDTIDQ 523

Db 480 LGLFALLPETKGLSLCDTME 501
RESULT 14
H86298
hypothetical protein AAD34689.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86298
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H86298
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <STO>
A:Cross-references: GB:AE005172; NID:g4966358; PIDN:AAD34689.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
Query Match 12.8%; Score 368.5; DB 2; Length 521;
Best Local Similarity 25.4%; Pred. No. 6.1e-21;
Matches 110; Conservative 94; Mismatches 180; Indels 49; Gaps 14;
QY 117 WEFSQDYVLTITVENLVCEDDWKAPLTISLFEVGVLLGSLGSLGOL-SDRCRKN-VLF 174
Db 98 WEWDGSGQKSVISEFDECCSSLLRGMPSFAFYIGAIVGGFFLALIPDDSLGRKKLVLF 157
QY 175 VTGMQTFGFSFLQIFSKNFEMFVFLVGVGMQISNVAAVFLVGLTEILG-----KSVRII 229
Db 158 STAMSI-TSIVISFTNWIYTFLEKFIIGFSRSQTSVALVLSERVSTRWRPRATMIP 216
QY 230 FSTLGVCFYAFGVYMLPLFAYFIR--WRMLVALTMPGV-LVALWVFPSPRWLS 286
Db 217 FT-----LFVLGMSLGLIAYLRQDSSWRYLYLTSVPAVFCIFLYLFALESPLMH 270
QY 287 QGRFEER-EVLIIRKAAKANGIVPSTIFDPSELQDLSKKQOSHNLID-LLRWNIRMT 344
Db 271 QGDKKEAIDVLTMSPKKAYL--ESVSKLPLKQENFEQAPYISINDFFRKAFFRIL 328
QY 345 IMSIMLWMTISVGYGLSLDTPNLHGDFVNCFLSAMVEVPAVYVLAWLLOLPRYSMA 404
Db 329 VVIMLF-GLGISYGYVPLAARDIDVNIYLSLSETLNALVELPTVITPILLERENRSSLV 387
QY 405 TALFLGGS--VLLFMQLVPPDLYLATVLMVGVK-----FGVTAASFMYVYT 450
Db 388 VNTLGGASGVLCF-----VLSILGKTEIAFAFELGTFCARIGNLMAVEN 434
QY 451 AELPTVVRNMGVSVSTASRLGSLTSPFYVYLGAIDRFLPYILMGLSLTILTAITLFLP 510
Db 435 VEMPTCVRSATMMLRQALVVGACCPLIASIGRYIPSVSFAIFGIMSGLGMFVLILP 494
QY 511 ESFGTPLPDTIDQ 523
Db 495 ETKGLSLCDSMEE 507
RESULT 15
T23190
hypothetical protein ZK637.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000
C:Accession: T23190
R:Craxton, M.

submitted to the EMBL Data Library, April 1993

A:Reference number: Z19704

A:Accession: T23190

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-529 <WIL>

A:Cross-references: EMBL:Z22175; PIDN:CAA80131.1; GSPDB:GN00021; CESP:ZK637.1

A:Experimental source: clone K01F9

C:Genetics:

A:Gene: CESP:ZK637.1

A:Map position: 3

A:Introns: 31/2; 121/3; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3

C:Superfamily: Caenorhabditis elegans glucose transport protein

```
Query Match      12.3%  Score 355.5;  DB 2;  Length 529;
Best Local Similarity 26.5%;  Pred. No. 6.3e-20;
Matches 140;  Conservative 97;  Mismatches 186;  Indels 105;  Gaps 24;

QY  52  VPDAANLSSANRNHTVPLRLRDGREVPHSCRRYRLATIANFSAIGLEP--GRDVD----- 104
Db  10  VLEASNLTEAVYDLTKAKLIKRIHVGGDF-AVRYSNLDDRTIELG-EPTDQRSPDSEKTF 67
QY  105 -----LCOLEQE--SCLDGWFEFSQD-----VYLSTI-----VTEWNLVCEDDWKAPLT 145
Db  68  TVDEAVEALGFRQLKLSILTGMAWADAMENMLLSLSPALACEWGI---SSVQQALV 124
QY  146  ISLPFVGLLSFGISQGLSDRFG-RKNVLFVT-----MGMTGFS---FLQIFSKNFEMP 196
Db  125  TTCVFSGMMLSTPWGKICDRGRKGLTFTLVACIMGVISGMSPHFYVLLFFRGLTGF 184
QY  197  VVFLVVGMGQISHVAAAFV--LGTEILGKSVRIESTLGVCFYAFGYMVLPLFAYFIR 254
Db  185 -----GIGGVPQSVTIYAEFLTAQRAKCVLIES-----FWAIGAVFEALLAYFVM 231
QY  255  D---WRMLLVALTNP-GVLCVALWPIPESPRLISQGRFEAEVIRKAAKANGIWPS 310
Db  232  ESFGWRALMFLSSGPLGIFAVAFW-LPESARFDMASGHPERALETLQAAARNRVOLPT 290
QY  311  TIFDPSELQDLSSKKQOSHNLDL--LRTWNIRMTVIMSIMLWMTISVGYFGLSLDTPNL 368
Db  291  G-----RLVSSIKAGSESGDIANLLSPDLRKTILLWCITAITAFSYGVMVLFYTVL 343
QY  369 -----HGDIFVN-----CFLSAMVEVPAYVLAWLLLOYLPRYSMA 404
Db  344  FQSHDECHGGLFSRGTMQEVCOPISTRSDYFDLLSTTLAEPPGLIITVLIIEWFGRKKTMA 403
QY  405  T--ALFLGGSVLLFMQLVPPDLYLATVLMVGKFGVTAAFSMVYVYTAELYPTVVRNMG 462
Db  404  LEYAVFAITFELLYFCLD----RTTVVLIFVARAFISGAFQCAVYVYTPVYPTTLRANG 459
QY  463  VGVSTASRLGSLISPYFVYLGAYDRFLPYILMGSLTILTAILTFLP 510
Db  460  LGTCSAMARIGAIVA-----SEKSLPIGIYGTAAIILGLIASLSLP 501
```

Search completed: August 16, 2001, 13:59:29

Job time: 156 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 14:04:00 ; Search time 18.15 Seconds
(without alignments)
1051.257 Million cell updates/sec

Title: US-09-521-195-3
Perfect score: 2883
Sequence: 1 MRDYDEVTAFLGNGFFORL.....HTRMLKQGERPTILKSTAF 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 2883 | 100.0 | 557 | 1 | OCN2_HUMAN |
| 2 | 2513 | 87.2 | 557 | 1 | OCN2_MOUSE |
| 3 | 2509 | 87.0 | 557 | 1 | OCN2_RAT |
| 4 | 387.5 | 13.4 | 751 | 1 | YX5_CAEEL |
| 5 | 357 | 12.4 | 1222 | 1 | YMP3_CAEEL |
| 6 | 355.5 | 12.3 | 529 | 1 | YU1_CAEEL |
| 7 | 308 | 10.7 | 445 | 1 | YGS_ECOLI |
| 8 | 291.5 | 10.1 | 435 | 1 | Y13_CAEEL |
| 9 | 287 | 10.0 | 400 | 1 | YCEI_BACSU |
| 10 | 267 | 9.3 | 443 | 1 | YAU_ECOLI |
| 11 | 265 | 9.2 | 459 | 1 | YDUK_ECOLI |
| 12 | 262.5 | 9.1 | 742 | 1 | SV2_RAT |
| 13 | 256 | 8.9 | 482 | 1 | YF1G_BACSU |
| 14 | 255.5 | 8.9 | 452 | 1 | YDJE_ECOLI |
| 15 | 252.5 | 8.8 | 457 | 1 | PCAK_ACICA |
| 16 | 250.5 | 8.7 | 472 | 1 | ARE_KLEOX |
| 17 | 243.5 | 8.4 | 551 | 1 | HGT1_KLEUL |
| 18 | 241.5 | 8.4 | 472 | 1 | ARAE_ECOLI |
| 19 | 240.5 | 8.3 | 495 | 1 | GTR3_CANFA |
| 20 | 239.5 | 8.3 | 451 | 1 | YVAT_BACSU |
| 21 | 239 | 8.3 | 461 | 1 | CSC_BACSU |
| 22 | 237.5 | 8.2 | 496 | 1 | GTR3_HUMAN |
| 23 | 234 | 8.1 | 494 | 1 | GTR3_SHEEP |
| 24 | 227 | 7.9 | 490 | 1 | GTR1_CHICK |
| 25 | 226.5 | 7.9 | 464 | 1 | GALP_ECOLI |
| 26 | 224.5 | 7.8 | 413 | 1 | MUCK_ACICA |
| 27 | 223 | 7.7 | 451 | 1 | GTR1_PIG |
| 28 | 222 | 7.7 | 763 | 1 | RG2_YEAST |
| 29 | 221 | 7.7 | 492 | 1 | GTR1_RAT |
| 30 | 220 | 7.6 | 492 | 1 | GTR1_BOVIN |
| 31 | 219.5 | 7.6 | 491 | 1 | XYLE_ECOLI |
| 32 | 219 | 7.6 | 493 | 1 | GTR3_MOUSE |
| 33 | 217 | 7.5 | 496 | 1 | GTR3_CHICK |

| | | | | | |
|----|-------|-----|-----|---|------------|
| 34 | 216 | 7.5 | 492 | 1 | GTR1_HUMAN |
| 35 | 216 | 7.5 | 492 | 1 | GTR1_MOUSE |
| 36 | 215 | 7.5 | 492 | 1 | GTR1_RABIT |
| 37 | 213.5 | 7.4 | 486 | 1 | YGR4_YEAST |
| 38 | 213 | 7.4 | 448 | 1 | PCAK_PSEPU |
| 39 | 213 | 7.4 | 818 | 1 | SNF3_YEAST |
| 40 | 208 | 7.2 | 493 | 1 | GTR3_RAT |
| 41 | 207.5 | 7.2 | 567 | 1 | HXT9_YEAST |
| 42 | 206.5 | 7.2 | 567 | 1 | HXTA_YEAST |
| 43 | 204 | 7.1 | 547 | 1 | GTR1_LEIDO |
| 44 | 201 | 7.0 | 592 | 1 | HXT5_YEAST |
| 45 | 200.5 | 7.0 | 522 | 1 | STP1_ARATH |

ALIGNMENTS

| | |
|--|--|
| RESULT 1 | |
| OCN2_HUMAN | |
| ID OCN2_HUMAN STANDARD; PRT; 557 AA. | |
| AC O76082; | |
| DT 01-OCT-2000 (Rel. 40, Created) | |
| DT 01-OCT-2000 (Rel. 40, Last annotation update) | |
| DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22, MEMBER 5), (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER). | |
| DE SLC22A5 OR OCTN2. | |
| GN Homo sapiens (Human). | |
| OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| OX NCBI_TaxID=9606; | |
| RN [1] | |
| RP SEQUENCE FROM N.A. | |
| RX MEDLINE=98289574; PubMed=9618255; | |
| RA Wu X., Prasad P.D., Leibach F.H., Ganapathy V.; | |
| RT "CDNA sequence, transport function, and genomic organization of human OCTN2, a new member of the organic cation transporter family."; | |
| RL Biochem. Biophys. Res. Commun. 246:589-595(1998). | |
| RN [2] | |
| RP SEQUENCE FROM N.A. | |
| RX TISSUE=Kidney; | |
| RC MEDLINE=98352077; PubMed=9685390; | |
| RA Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M., | |
| Sai Y., Tsuji A.; | |
| RT "Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2."; | |
| RL J. Biol. Chem. 273:20378-20382(1998). | |
| RN [3] | |
| RP SEQUENCE FROM N.A. | |
| RX MEDLINE=99113835; PubMed=9916797; | |
| RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N., | |
| Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsulishi T., | |
| Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M., | |
| Tsuji A.; | |
| RT "Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter."; | |
| RL Nat. Genet. 21:91-94(1999). | |
| RN [4] | |
| RP CHARACTERIZATION. | |
| RA MEDLINE=99384224; PubMed=10454528; | |
| RX Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H., | |
| Chen J., Conway S.J., Ganapathy V.; | |
| RT "Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter."; | |
| RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999). | |
| RN [5] | |
| RP VARIANT CDSP GLN-169. | |
| RX MEDLINE=99355597; PubMed=10425211; | |
| RA Burwinkel B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K., | |
| Gerbitz K.-D., Killmann M.W.; | |
| RT "Carnitine transporter OCTN2 mutations in systemic primary carnitine deficiency: a novel Arg169Gln mutation and a recurrent Arg282ter | |

mutation associated with an unconventional splicing abnormality.";
 Biochem. Biophys. Res. Commun. 261:484-487(1999).
 [6] VARIANTS CDSP CYS-211.
 MEDLINE=99408248; PubMed=10480371;
 RA Vaz F.N., Scholte H.R., Ruiter J., Huisaarts-Odijk L.M.,
 RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,
 RA Wanders R.J.A.;
 RT "Identification of two novel mutations in OCTN2 of three patients with
 RT systemic carnitine deficiency.";
 RT Hum. Genet. 105:157-161(1999).
 [7] VARIANTS CDSP LEU-478.
 MEDLINE=99172075; PubMed=10072434;
 RA Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,
 RA Wanders R.J., Fok T.F., Hjeltnen N.M.;
 RA "Mutations of OCTN2, an organic cation/carnitine transporter, lead to
 RT deficient cellular carnitine uptake in primary carnitine deficiency.";
 RT Hum. Mol. Genet. 8:655-660(1999).
 [8] CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.
 MEDLINE=20026865; PubMed=1059218;
 RA Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;
 RA "Mutations in novel organic cation transporter (OCTN2), an organic
 RT cation/carnitine transporter, with differential effects on the
 RT organic cation transport function and the carnitine transport
 RT function.";
 RT J. Biol. Chem. 274:33388-33392(1999).
 [9] VARIANTS CDSP ARG-383 AND PHE-446.
 MEDLINE=20081068; PubMed=10612840;
 RA Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,
 RA Tsuji A.;
 RA "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a
 RT patient with primary systemic carnitine deficiency.";
 RT Hum. Mutat. 15:118-118(2000).
 [10] VARIANTS CDSP LYS-452.
 MEDLINE=20145665; PubMed=10679939;
 RA Wang Y., Kelly M.A., Cowan T.M., Longo N.;
 RA "A missense mutation in the OCTN2 gene associated with residual
 RT carnitine transport activity.";
 RT Hum. Mutat. 15:238-245(2000).
 CC -!- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
 CC CARNITINE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,
 CC HEART AND PLACENTA.
 CC -!- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY
 CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE
 CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE
 CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND
 CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL
 CC MYOPATHY OR CARDIOMYOPATHY.
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
 CC CATION SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF057164; AAC24828.1; -;
 DR EMBL; AB015050; BAA29023.1; -;
 DR EMBL; AB016625; BAA36712.1; -;
 DR MIM; 603377; -;
 DR MIM; 212140; -;
 DR InterPro; IPR001066; -;

DR InterPro; IPR001687; -;
 DR Pfam; PF00083; sugar_tr; 1;
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1. Disease mutation.
 KW Transport; Transmembrane; Glycoprotein; POTENTIAL.
 FT TRANSMEM 21 41
 FT TRANSMEM 143 163
 FT TRANSMEM 173 193
 FT TRANSMEM 196 216
 FT TRANSMEM 233 253
 FT TRANSMEM 258 278
 FT TRANSMEM 343 363
 FT TRANSMEM 372 392
 FT TRANSMEM 414 434
 FT TRANSMEM 437 457
 FT TRANSMEM 489 509
 FT CARBOHYD 57 57
 FT CARBOHYD 64 64
 FT CARBOHYD 91 91
 FT VARIANT 169 169
 FT VARIANT 211 211
 FT VARIANT 283 283
 FT VARIANT 446 446
 FT VARIANT 452 452
 FT VARIANT 478 478
 FT MUTAGEN 352 352
 FT SEQUENCE 557 AA; 62751 MW; 928B1F6EF63C48D CRC64;
 SQ
 Query Match 100.0%; Score 2883; DB 1; Length 557;
 Best Local Similarity 100.0%; Pred. No. 3.1e-178;
 Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRDDEVTAFLGEGWPPQRLIFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAAANLS 60
 DB 1 MRDDEVTAFLGEGWPPQRLIFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAAANLS 60
 QY 61 AWRNHTVPLRLRDGREVPHSCRRYRLATIANFSALEPGRDVLQGLEQESCLDGWEFS 120
 DB 61 AWRNHTVPLRLRDGREVPHSCRRYRLATIANFSALEPGRDVLQGLEQESCLDGWEFS 120
 QY 121 QDVYLSITVTENLVCEDDWKAPLTISLFPVGLLGSFISGQLSDRGRKNVLFVTMGWQ 180
 DB 121 QDVYLSITVTENLVCEDDWKAPLTISLFPVGLLGSFISGQLSDRGRKNVLFVTMGWQ 180
 QY 181 TGFSEPLQIFSKNFEMFVFLVVLGGMGOISNVAAFLVGLTGLKSVRIIFSTLGVCFYFA 240
 DB 181 TGFSEPLQIFSKNFEMFVFLVVLGGMGOISNVAAFLVGLTGLKSVRIIFSTLGVCFYFA 240
 QY 241 FGVMVPLFATFIRDWRMLVALTMPGVLCVLMWFTIPSPRWLISQGRFEEAEVIRKA 300
 DB 241 FGVMVPLFATFIRDWRMLVALTMPGVLCVLMWFTIPSPRWLISQGRFEEAEVIRKA 300
 QY 301 AKANGIVVPSTIFDPSELDLSSKKQKSHNLLDLRTWNIRMTVIMSIMLWMTISVGYFG 360
 DB 301 AKANGIVVPSTIFDPSELDLSSKKQKSHNLLDLRTWNIRMTVIMSIMLWMTISVGYFG 360
 QY 361 LSLDTPNLHGDIFFVNCFLSAMVEVPAYVLAWLQLQYLPRIYSMATALFLGGSVLLPMQLV 420
 DB 361 LSLDTPNLHGDIFFVNCFLSAMVEVPAYVLAWLQLQYLPRIYSMATALFLGGSVLLPMQLV 420
 QY 421 PDLVYLATVLMVWVKFGVTAAFSMVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 480
 DB 421 PDLVYLATVLMVWVKFGVTAAFSMVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 480

Qy 481 VYLGYDRFLPYILMGSLTILTAITLFLPESFGTPLPDTIDQMLRVKGMKHKRTPSHTR 540
 Db 481 VYLGYDRFLPYILMGSLTILTAITLFLPESFGTPLPDTIDQMLRVKGMKHKRTPSHTR 540
 Qy 541 MKDQGERPTILKSTAF 557
 Db 541 MKDQGERPTILKSTAF 557

RESULT 2

OCN2_MOUSE
 ID OCN2_MOUSE STANDARD; PRT; 557 AA.
 AC Q920E8;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
 MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).
 GN SLC22A5 OR OCTN2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RX MEDLINE=99113835; PubMed=9916797;
 RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,
 RA Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsushita T.,
 RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,
 RA Tsuji A.;
 RT "Primary systemic carnitine deficiency is caused by mutations in a
 RT gene encoding sodium ion-dependent carnitine transporter.";
 RL Nat. Genet. 21:91-94(1999).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.
 RC STRAIN=C3H;
 RX MEDLINE=99057546; PubMed=9837751;
 RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;
 RT "A missense mutation of mouse OCN2, a sodium-dependent carnitine
 RT cotransporter, in the juvenile visceral steatosis mouse.";
 RL Biochem. Biophys. Res. Commun. 252:590-594(1998).
 RN [3]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=99384224; PubMed=10454528;
 RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
 RA Chen J., Conway S.J., Ganapathy V.;
 RT "Functional characteristics and tissue distribution pattern of organic
 RT cation transporter 2 (OCTN2), an organic cation/carnitine
 RT transporter.";
 RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
 CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
 CC CARNITINE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL
 CC STEATOSIS (JVS).
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
 CC CATION SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB015800; BAA36590.1; -;
 DR EMBL; AF111425; AAC99787.1; -;
 DR EMBL; AF110417; AAD54060.1; -;
 DR InterPro; IPR001066; -;

DR Pfam; PF00083; sugar_tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
 KW Transport; Transmembrane; Glycoprotein; Disease mutation.
 FT TRANSMEM 21 41
 FT TRANSMEM 143 163
 FT TRANSMEM 173 193
 FT TRANSMEM 198 218
 FT TRANSMEM 233 253
 FT TRANSMEM 258 278
 FT TRANSMEM 342 362
 FT TRANSMEM 374 394
 FT TRANSMEM 407 427
 FT TRANSMEM 431 451
 FT TRANSMEM 489 509
 FT CARBOHYD 57 57
 FT CARBOHYD 64 64
 FT CARBOHYD 91 91
 FT CARBOHYD 322 322
 FT VARIANT 352 352
 SQ SEQUENCE 557 AA; 62779 MW; 6093F0EE9612B204 CRC64;

Query Match 87.2%; Score 2513; DB 1; Length 557;
 Best Local Similarity 85.5%; Pred No. 1.7e-154;
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;
 Qy 1 MRDYDEVTAFLGEGWGPQRLIFELLSSAIIIPNGFTGLSSVFLIATPEHRCRVPDAAANLSS 60
 Db 1 MRDYDEVTAFLGEGWGPQRLIFELLSSAIIIPNGFTGLSSVFLIATPEHRCRVPDAAANLSS 60
 Qy 61 AWRNHTVPLRDGREGVPHSCRRYRLATIANFSALGLEPGRDVDLQLEQESCLDGEWERS 120
 Db 61 AWRNHSIPLTKDGRQVQPKRRYRLATIANFSELGLEPGRDVDLQLEQESCLDGEWYD 120
 Qy 121 ODVYLSITVTEWNLVCEDDMKAPLITSLFEVGLVLSGSLSDRFRGNKLVLFVTGMQ 180
 Db 121 KDVLSTIVTEWDLVCKDDMKAPLITSLFEVGLVLSGSLSDRFRGNKLVLFVTGMQ 180
 Qy 181 TGSFLOIFSKNFEMFVFLVFLVGMGQISNYVAAFVLGTILGSKSVRIIFSTLGVCFYA 240
 Db 181 TGSFLOVFSVNFEMFVFLVFLVGMGQISNYVAAFVLGTILGSKSVRIIFATLGVCFYA 240
 Qy 241 FGYMVLPLFAYFIRDRMLLVALTMPGVLCALWFWFIPESPRWLLISQGRKEAEVIIRKA 300
 Db 241 FGYMVLPLFAYFIRDRMLLVALTMPGVLCALWFWFIPESPRWLLISQGRKEAEVIIRKA 300
 Qy 301 AKANGIVVPSTIFDPSELQDLSKQKQSHNLDLRTNIRMTVIMSLMWTISVGYFG 360
 Db 301 AKINGIVAPSTIFDPSELQDLSKQKQSHNLDLRTNIRMTVIMSLMWTISVGYFG 360
 Qy 361 LSLDTPNLHGDIYVNCFLSAMVEVPAYVLAWLLIQLYLPFRYSMATALLFGSGVLLFMQLV 420
 Db 361 LSLDTPNLHGDIYVNCFLSAMVEVPAYVLAWLLIQLYLPFRYSMATALLFGSGVLLFMQLV 420
 Qy 421 PPDLYLATLVLMVGKFGVTAFAFVVMVYVYAEVPTVVRNMGVGVSTASRLSILSPYF 480
 Db 421 PSELFYLTALVMVGKFGVTAFAFVVMVYVYAEVPTVVRNMGVGVSTASRLSILSPYF 480
 Qy 481 VYLGYDRFLPYILMGSLTILTAITLFLPESFGTPLPDTIDQMLRVKGMKHKRTPSHTR 540
 Db 481 VYLGYDRFLPYILMGSLTILTAITLFLPESFGTPLPDTIDQMLRVKGMKHKRTPSHTR 540
 Qy 541 MKDQGERPTILKSTAF 557
 Db 541 MKDQGERPTILKSTAF 557
 RESULT 3
 OCN2_MOUSE STANDARD; PRT; 557 AA.
 AC O70594; Q9QW0L;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)

01-OCT-2000 (Rel. 40, Last annotation update)
ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER)
(UST2R) (CTI).
GN SLC22A5 OR OCTN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Kidney;
RX MEDLINE=98200080; PubMed=9541011;
RA Schoenig E., Spitzberger F., Engelhardt M., Martel F., Oerding N.,
Grudemann D.;
RT "Molecular cloning and characterization of two novel transport
proteins from rat kidney.";
RL FEBS Lett. 425:79-86(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;
RX MEDLINE=99011422; PubMed=9792817;
RA Sekine T., Kusunahara H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,
Kanai Y., Endou H.;
RT "Molecular cloning and characterization of high-affinity carnitine
transporter from rat intestine.";
RL Biochem. Biophys. Res. Commun. 251:586-591(1998).
[3]
RN SEQUENCE FROM N.A.; AND CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=1045428;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
cation transporter 2 (OCTN2), an organic cation/carnitine
transporter.";
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CARNITINE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES
AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND
ARTERIOLES IN THE HEART, IN THE LABYRINTHINE LAYER OF THE
PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE
BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CATION SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AJ001933; CAA05106.1; -
EMBL; AB017260; BAA34399.1; -
EMBL; AF110416; AAD54059.1; -
InterPro; IPR001066; -
Pfam; PF00083; sugar_tr. 1.
PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 342 362 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.

FT TRANSMEM 431 451 POTENTIAL.
FT TRANSMEM 489 509 POTENTIAL.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 275 275 W -> G (IN REF. 2).
SQ SEQUENCE 557 AA; 62567 MW; 172472E7B0A5F043 CRC64;

Query Match 87.0%; Score 2509; DB 1; Length 557;
Best Local Similarity 85.5%; Pred. No. 3.le-154;
Matches 476; Conservative 42; Mismatches 39; Indels 0; Gaps 0;

QY 1 MRDYDEVTATLGEWGPQRLIFELLSASIPNGFTGLSSVFLIATPEHRCRVEDAANLS 60
DB 1 MRDYDEVTATLGEWGPQRLIFELLSASIPNGFTGLSSVFLIATPEHRCRVEDAANLS 60

QY 61 AWRNHTVPLRLDRGVRPHSCRRYRLATIANFSALGLEPRDVLGQLEQSCLDGWFEF 120
DB 61 AWRNHSIPLETKDGRVQPSQRRYRLATIANFSALGLEPRDVLGQLEQSCLDGWFEY 120

QY 121 QDYVLTIVTEWNLVCEDDWKAPLTISLFFVGLGSLGSGSDFGRKKNVLFVTMGWQ 180
DB 121 KDVFSLTIVTEWDLVCKDKWKAPLTISLFFVGLGSLGSGSDFGRKKNVLFVTMGWQ 180

QY 181 TGFSLQIFSKNEMFVFLVVGMOISNYVAAFVLGTEILGKSVRIIFSTLGVCIFFA 240
DB 181 TGFSLQLFSVNEFMTVLVVGMOISNYVAAFVLGTEILGKSVRIIFATLGVCIFFA 240

QY 241 FGMVPLPLFAFTRDWRMLLVALTMPGVLGVCLVWLPFIPSPRWLISQGRFEAEVIRKA 300
DB 241 FGMVPLPLFAFTRDWRMLLVALTMPGVLGVCLVWLPFIPSPRWLISQGRFEAEVIRKA 300

QY 301 AKANGIVPSTIEDPSELQDLSSKKOOSHNLDLRTNMRVMTIMSLMWTISVGYFG 360
DB 301 AKNGIVAPSTIEDPSELQDLSSKKPQSHIYDLVTRNIRITIMSLILWLTISVGYFG 360

QY 361 LSLDTPMLHGDIIVNCFLSAMVEVPAYVLAAILLQYLPYRYMATLFGGSLVFLMOLY 420
DB 361 LSLDTPMLHGDIIVNCFLLAAVEVPAYVLAAILLQYLPYRYMATLFGGSLVFLIOLV 420

QY 421 PDLVLYATLVLMVVGKFGVTAFAFVSVVYTAELYPTVVRNMGVSVSTASRLGSLSPYF 480
DB 421 PSELYLSTALVVMVGFGITSAISVSVVYTAELYPTVVRNMGVSVSTASRLGSLSPYF 480

QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKMGKHKRTPSHR 540
DB 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKMGKHKRTPSHR 540

QY 541 MKDQGERPTILKSTAF 557
DB 541 TQKDGGSPTVLKSTAF 557

RESULT 4
YLX5_CAEEL STANDARD; PRT; 751 AA.
ID YLX5_CAEEL
AC P46501;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL 84.8 KDA PROTEIN F23F12.5 IN CHROMOSOME III.
GN F23F12.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du 2;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U12965; AAA20607.1; -
WormPep; F23F12.5; CE01252.
InterPro; IPR001066; -
Pfam; PF00083; sugar tr; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 102 122 POTENTIAL.
FT TRANSMEM 137 157 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 320 340 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 410 430 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT TRANSMEM 515 535 POTENTIAL.
FT TRANSMEM 547 567 POTENTIAL.
FT TRANSMEM 583 603 POTENTIAL.
FT TRANSMEM 614 634 POTENTIAL.
FT TRANSMEM 678 698 POTENTIAL.
SQ SEQUENCE 751 AA; 84832 MW; A6C4F43540295EFC CRC64;

Query Match 13.4%; Score 387.5; DB 1; Length 751;
Best Local Similarity 25.8%; Pred. No. 6.3e-18;
Matches 106; Conservative 89; Mismatches 189; Indels 27; Gaps 9;

QY 127 TIVTEWNLVC-EDDWKAPLTISLFFVGLGSPISQLSDFRGKKNVLFVTMGQTGFSF 185
DB 303 SVMDFKMFCTKAYDAWVATQIFGLVIGAITYGHLDHFGKRPVSFFGISVGLFGV 362
QY 186 LQIFSKNFEMFVFLVFLVGMGQISNYVAAPVFLGTEILGKSVRIIFSTLGVCFIYAPGY-- 243
DB 363 ASGFAPSWVEFAAFREIVGFSIASILIVFAYILEIEPEQVFLAS-----FFNNGYAR 417
QY 244 MVLPFAFYTRDRMLLVATMPGVLCVLAWMFIPSPRWLISQGRFEAEVIRKAAGA 303
DB 418 LVFTLACFCGYWRSAAIATSLSLPLVILLIPSPKWNFKKFRDARAARAVML 477
QY 304 NGIVVPSTIFDPSELQD-LSSKKQOSHNLDLRTWNRIMVTIMSLMWTISVGYEGLS 362
DB 478 SGPIYVNDQDSIEISKLEEKSTKIYTKDLFTSWTIAIYIVGSLMFSTLSAFGSD 537
QY 363 LDTPNLHGDIFVACFLSAMVEVPAYVLAWLLOLPLP-----RRYSM--ATAFLFGGSVLLF 416
DB 538 LNSGNLAGNYLSQFVSGVATFAKIFVFLDITYVDFDRRLHQYQJLAMILCYCIVN 597
QY 417 MLQVPP-----DLYYIATVLMVVGKFGVTAASVMVYVTAELYPTVVRNMGVGS 467
DB 598 LMILPESDCSGSQRDLAIT--IINIIGVSFIETWDCVLAIVECFPTKIRTIGTGCS 655
QY 468 TASRLGSILSPYFVYLGAYDRFLPYILMGLS-LTILTAITLFLPESFGTPL 517
DB 656 LLARTGALLAPQAWYLSDIIRPAPYAVCVCSIGTISLLISCVFPLPDKGVDL 706

RESULT 5
YMP3_CAEL
ID YMP3_CAEL STANDARD; PRT; 1222 AA.
AC Q10947;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE HYPOTHETICAL 139.9 KDA PROTEIN B0361.3 IN CHROMOSOME III.
GN B0361.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00031; AAA50622.1; -
DR WormPep; B0361.3; CE00752.
DR Pfam; PF00083; sugar tr; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 232 252 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 307 327 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 355 375 POTENTIAL.
FT TRANSMEM 382 402 POTENTIAL.
FT TRANSMEM 413 433 POTENTIAL.
FT TRANSMEM 442 462 POTENTIAL.
SQ SEQUENCE 1222 AA; 139868 MW; CBA42A80A254FB80 CRC64;

Query Match 12.4%; Score 357; DB 1; Length 1222;
Best Local Similarity 21.1%; Pred. No. 9.4e-16;
Matches 118; Conservative 106; Mismatches 209; Indels 126; Gaps 14;

QY 3 DYDEVTAFLEGGPPQRIFFLLSASIPNGFTGLSSVFLIA---TPEHRCRVPDAANL 58
DB 35 DPKDEVEAYGAYGKYQIIFYVLVQPL----NFFYSSSMYIMSFVOLNLEKOCE----- 83
QY 59 SSAMENHTVPLRLDRGVPHSCRRYRLATIANFSALEPGRDVDLGLQEQESCLOGWE 118
DB 84 ---YKNEIP-----ISETCQ-----IETESKAFGNLNGEYC---G 114
QY 119 FSQDYVL-----STIVTEWNLVCEDDWKAPLTISLFFVGLGSPISQLSDFRGK 170
DB 115 TAENTLVNVTNQKASTNLLVDPLDLSCHSHWFQEFGLTIFTIGAVIAVPFMSMLADRYGR 174
QY 171 NVLFVTMGQGTGFSLQIFSKNFEMFVFLVFLVGMGQISNYVAAPVFLGTEILGKSVRIIF 230
DB 175 PLIVTTAILAFLANMAASFSNFAIFLILRAFIGACSDSYLSVASVATCEYLSEKAR-AW 233
QY 231 STLGVCIFFVAFGYMVLPLFAVFIWRMLLVATMPGVLCVLAWMFIPESPRWLISQGRF 290
DB 234 ITVVYVNAWSLGMVWVTLVLTMTDWRWRYFVLSLPGVYGVALWFFPESPHLLTKNKT 293
QY 291 EBAEVIIRKAAGKANGIVVPSTIFDPSELQDLSKKQOSHNLDLRTWNRIMVTIMSL 350
DB 294 EKLKKYIKTANR-----
QY 351 WMTISVGYEGLSLDTPNLHGDIFVNCFL-SAMVEVPAYVLAWLLOLPLPRYSM----- 403
DB 306 -NVISLVYFAISFMSVELGGD-QVQAFYLSLIEIPAGLVIPLMMKMKRKNIVICLVF 363
QY 404 ATALFLGGSVLLFMQLVPPDLYIATVLMVVGKFGVTAASVMVYVTAELYPTVVRNMGV 463

Db 364 QTLALIGTVFL-----DSVEFKVIMLVAKVMATIIYSVHPDWATEQPTSVRSICF 416

QY 464 GVSSTASRLGSLSPYVYLGAYDFLPYILMGSLTILTAITLFLPESFGPLPDTIDQ 523

Db 417 SLMNIPQSMGIIMSPVVKHIVSNPWIPVVLTFISATLAEMLHETKNKLLPTDIES 476

QY 524 MLRVGKMKHKRTPSHRML 542

Db 477 L-----SYPSSETNDL 486

RESULT 6

YOUNL_CAEEL STANDARD; PRT; 529 AA.

AC P30638; Q21101;

DT 01-APR-1993 (Rel. 25, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III.

GN ZK637.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=92168156; PubMed=1538779;

RA Sulston J., Du Z., Thomas K., Willner R., Hillier L., Staden R.,

RA Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,

RA Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,

RA Ainscough R., Waterston R.;

RT "The C. elegans genome sequencing project: a beginning.";

RL Nature 356:37-41(1992).

RN [2]

RP REVISIONS.

RC STRAIN-BRISTOL N2;

RA Durbin R.;

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).

CC -|- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z11115; CAA77460.1; -

DR EMBL; Z22175; CAA77460.1; JOINED.

DR EMBL; Z22175; CAA80131.1; -

DR EMBL; Z11115; CAA80131.1; JOINED.

DR PIR; S15786; S15786.

DR WormPep; ZK637.1; CE06638.

DR InterPro; IPR001066; -

DR Pfam; PF00083; sugar_tk; 1

DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.

DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.

KW Hypothetical protein; Duplication; Transmembrane; Transport.

FT TRANSMEM 86 106 POTENTIAL.

FT TRANSMEM 122 142 POTENTIAL.

FT TRANSMEM 158 178 POTENTIAL.

FT TRANSMEM 238 258 POTENTIAL.

FT TRANSMEM 320 340 POTENTIAL.

FT TRANSMEM 373 393 POTENTIAL.

FT TRANSMEM 411 431 POTENTIAL.

FT TRANSMEM 482 502 POTENTIAL.

SQ SEQUENCE 529 AA; 58317 MW; 8D2FF4CBA15ECD2D CRC64;

Query Match 12.3%; Score 355.5; DB 1; Length 529;

Best Local Similarity 26.5%; Pred. No. 5e-16;

Matches 140; Conservative 97; Mismatches 186; Indels 105; Gaps 24;

QY 52 VPDAAANLSSAWRNHTVPLRLRDGREVPHSCRRYRLATIANFSAIGLEP--GROVD----- 104

Db 10 VLEASNLTEAYVLTAKQLKEIRHVGDDP-AVRYSNLDDORTELG-EPTDQRSPDSEKTF 67

QY 105 -----LGQLEQF-SCLDGWESQD-----VYLSTI-----VTEWNLVCEDDWKAPLT 145

Db 68 TVDEAVEALGGRFQLKLSILTGMANMADAMEMMLLSLSPALACEWGI---SSVQQALV 124

QY 146 ISLEFFVGLGSPFISGQSDRFG-RKNVLVFT-----MGMTGFS---FLQIFSKNFMFP 196

Db 125 TTCVFSGMMLSSTFWGXKICDRFGRKGLTFTLVACIMGVISGMSPHFYVLLFRGLTGF 184

QY 197 VLVFLVLMGQISNYVAAFV--LGTEILGKSVRIESTLGVCFYAFGYMVLPLFAFYFIR 254

Db 185 -----GIGVFQSVLYIAEFLPTQRAKCVVLIIS-----FWAIGAVFEALLAYFVM 231

QY 255 D---WRMLLVALTMP-GVLCAVLMWFIPESPRWLISQGRFEEAEVIRKAAKANGIVVPS 310

Db 232 ESFGWRALMFLSLPLGIFAVASFW-LPESARFDMASGHPERALETLQAAARNRVQLPT 290

QY 311 TIFDPSELQDLSSKKQOSHNLIDL--LRTWNIRMVIMSLMWTISVGFGSLDTPNL 368

Db 291 G-----RLVSTKAGSESRGDIANLLSPDLRKTILLWCIIWATAFSYGMVLFETVL 343

QY 369 -----HGDIFFVN-----CFLSAMVEVPAYVLAWLILLOYLPRRYSMA 404

Db 344 FQSHDECHGGLFNSGTQMEVCQPLTRSDYFDLLSTTLAEPPGLIITVLIIEWGRKKTMA 403

QY 405 T--ALFLGGVSLFMQLVPPDLYLATVLMVGKFGVTAFAFVSVYVYTAELYPTVYVRNMG 462

Db 404 LEVAVFAIFTFLLYFCLD----RFTVTVLIFVARAFISGAFOCAVYVTPDEVYPTTLRAVG 459

QY 463 VGVSTASRLGSLSPYVYLGAYDFLPYILMGSLTILTAITLFLP 510

Db 460 LGTCSAMARIGAIVA-----SEKSLPIGIYGTAILGLIASLIP 501

RESULT 7

YGCS_ECOLI STANDARD; PRT; 445 AA.

AC Q46909;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN CYSJ-ENO INTERGENIC

DE REGION.

GN YGCS.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MGL1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE

CC (POTENTIAL).

CC -|- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
DR EMBL; U29579; AAA69281.1; ALT_INIT.
DR EMBL; AE000360; AAC75813.1; ALT_INIT.
DR EcoGene; EG13126; ygcS.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00216; SUGAR_TRANSPORT_2; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transmembrane; Inner membrane.
FT TRANSMEM 23 43
FT TRANSMEM 33 43
FT TRANSMEM 57 77
FT TRANSMEM 86 106
FT TRANSMEM 115 135
FT TRANSMEM 143 163
FT TRANSMEM 176 196
FT TRANSMEM 254 274
FT TRANSMEM 287 307
FT TRANSMEM 312 332
FT TRANSMEM 338 358
FT TRANSMEM 370 390
FT TRANSMEM 401 421
FT TRANSMEM 445 AA; 48234 MW; B59E452721B15774 CRC64;
SQ SEQUENCE 445 AA; 48234 MW; B59E452721B15774 CRC64;

Query Match 10.7%; Score 308; DB 1; Length 445;
Best Local Similarity 26.6%; Pred. No. 4.7e-13;
Matches 102; Conservative 79; Mismatches 168; Indels 34; Gaps 11;

QY 151 VGVLLSFTSGQLSDRFGRKNVLFVTMGMTGTFSLQIFSKNFEMFVFLVFLVGMQISN 210
DB 151 VGVLLSFTSGQLSDRFGRKNVLFVTMGMTGTFSLQIFSKNFEMFVFLVFLVGMQISN 210
QY 65 LGLFLGSLVGLWISDHRGKQIFTSFLLITLASFLQFFATPEHLIGRLILIGLGGD 124
DB 65 LGLFLGSLVGLWISDHRGKQIFTSFLLITLASFLQFFATPEHLIGRLILIGLGGD 124
QY 211 YVAAFVLGTIELKSVRIIFSTLGV-CIFYAFGYMVLPLFA-YFIRD----WRMLLVALT 264
DB 211 YVAAFVLGTIELKSVRIIFSTLGV-CIFYAFGYMVLPLFA-YFIRD----WRMLLVALT 264
QY 125 YSVCHTLLAEFSRRHGIL--LGAHSVVTWYGVVLASITAGHHFISEPNEARWLLASAA 182
DB 125 YSVCHTLLAEFSRRHGIL--LGAHSVVTWYGVVLASITAGHHFISEPNEARWLLASAA 182
QY 265 MPGVLCVALMWFIPESPRWLISGRFEAEVIRKAAKANGIVVPSTIPDPSEL--QDLS 322
DB 265 MPGVLCVALMWFIPESPRWLISGRFEAEVIRKAAKANGIVVPSTIPDPSEL--QDLS 322
QY 183 LPALLITLLNWGTPSPRWLLRQGRFAEAHAIVHR-----YEGPHVLLGDEVV 230
DB 183 LPALLITLLNWGTPSPRWLLRQGRFAEAHAIVHR-----YEGPHVLLGDEVV 230
QY 323 SKKQSHNILLRTNIRMTVIMSIMLWMTISVGYEGLSDLPNLHGDFVNCFLSAMV 382
DB 323 SKKQSHNILLRTNIRMTVIMSIMLWMTISVGYEGLSDLPNLHGDFVNCFLSAMV 382
QY 231 TATHKHITLFFSRYWR---RTAFNSVFEVCLVPIWFIYTWLPTAQITGLDALTASL 287
DB 231 TATHKHITLFFSRYWR---RTAFNSVFEVCLVPIWFIYTWLPTAQITGLDALTASL 287
QY 383 EVPAYVLAWLILLYLPRLRYSMATLFLGGSVLIF-----MQLVPPDLYLATVLMVGKF 437
DB 383 EVPAYVLAWLILLYLPRLRYSMATLFLGGSVLIF-----MQLVPPDLYLATVLMVGKF 437
QY 288 MNNALLIVGALLG-LVLTLLAHRKELGSLFLLAATLVMACLPSGSSITLLLVLFST 346
DB 288 MNNALLIVGALLG-LVLTLLAHRKELGSLFLLAATLVMACLPSGSSITLLLVLFST 346
QY 438 GVTAAFSVMYVYTAELYPTVVRNMGVSVSTASRLGSLSPYV--YLGAYDRFLPYILM 495
DB 438 GVTAAFSVMYVYTAELYPTVVRNMGVSVSTASRLGSLSPYV--YLGAYDRFLPYILM 495
QY 347 TISAVSNLVGLPAESFPPTDIRSLGVGFATMSRLGAASVSTGLLPWVLAQWGMQVTLILL 406
DB 347 TISAVSNLVGLPAESFPPTDIRSLGVGFATMSRLGAASVSTGLLPWVLAQWGMQVTLILL 406
QY 496 GSLTILTAIT-LFLPESFGTPL 517
DB 496 GSLTILTAIT-LFLPESFGTPL 517
QY 407 ATVLLGVVVTWLPWAPETKALPL 429
DB 407 ATVLLGVVVTWLPWAPETKALPL 429

RESULT 8
YTI3_CAEL
ID YTI3_CAEL STANDARD; PRT; 435 AA.
AC Q10917;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 48.6 KDA PROTEIN B0252.3 IN CHROMOSOME II.
GN B0252.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]

SEQUENCE FROM N.A.
RP YCEI_BACSU
ID YCEI_BACSU STANDARD; PRT; 400 AA.
AC O34691;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN RAPJ-OPUAA INTERGENIC
DE REGION.
GN YCEI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.

SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RA Du 2., Waterston R.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


Query Match 9.2%; Score 265; DB 1; Length 459;
Best Local Similarity 26.2%; Pred. No. 2.8e-10;
Matches 101; Conservative 79; Mismatches 153; Indels 52; Gaps 18;


```
RESULT 12
SV2_RAT
ID SV2_RAT STANDARD; PRT; 742 AA.
AC Q02563;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SYNAPTIC VESICLE PROTEIN 2 (SV2).
GN SV2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 1-40.
RC TISSUE=Brain;
RX MEDLINE=92390722; PubMed=1519064;
RA Bajjalieh S.M., Peterson K., Shingal R., Scheller R.H.;
RT "SV2, a brain synaptic vesicle protein homologous to bacterial
transporters";
RL Science 257:1271-1273(1992).
CC -1- FUNCTION: MAY ACT AS A VESICLE-LOCALIZED NEUROTRANSMITTER
TRANSPORTER OR MAY FUNCTION AS AN ION TRANSPORTER OR CHANNEL.
CC -1- SUBCELLULAR LOCATION: SYNAPTIC VESICLE.
CC -1- TISSUE SPECIFICITY: NEURAL AND ENDOCRINE CELLS OF BRAIN AND
SPINAL CORD.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
CC
DR EMBL; L05435; AAA42188.1; -.
KW Pfam; P00083; sugar_tr; 1.
DR Synapse; Nerve; Glycoprotein; Neurotransmitter transport;
Transmembrane.
FT DOMAIN 1 163 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 164 182 POTENTIAL.
FT DOMAIN 183 204 LUMENAL (POTENTIAL).
FT TRANSMEM 205 225 POTENTIAL.
FT DOMAIN 226 242 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 243 260 POTENTIAL.
FT DOMAIN 261 262 LUMENAL (POTENTIAL).
FT TRANSMEM 263 281 POTENTIAL.
FT DOMAIN 282 294 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 295 322 POTENTIAL.
FT DOMAIN 323 334 LUMENAL (POTENTIAL).
FT TRANSMEM 335 355 POTENTIAL.
FT DOMAIN 356 445 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 446 469 POTENTIAL.
FT DOMAIN 470 592 LUMENAL (POTENTIAL).
FT TRANSMEM 593 611 POTENTIAL.
FT DOMAIN 612 626 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 627 647 POTENTIAL.
FT DOMAIN 648 649 LUMENAL (POTENTIAL).
FT TRANSMEM 650 669 POTENTIAL.
FT DOMAIN 670 694 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 695 711 POTENTIAL.
FT DOMAIN 712 712 LUMENAL (POTENTIAL).
FT TRANSMEM 713 731 POTENTIAL.
FT DOMAIN 732 742 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 742 AA; 82705 MW; 565DE7EF2929D5B CRC64;
```

Query Match

9.18; Score 262.5; DB 1; Length 742;

Best Local Similarity

17.5%; Pred. No. 6.6e-10;

```
Matches 115; Conservative 103; Mismatches 173; Indels 267; Gaps 21;
QY 96 GLEPGRDVLDGQLEQESCLDGEFSODVYLSTIVTE-----WNL----- 134
DB 125 GLSDGEGPPGGRGAORRKRDELAQO--YETILRCGHRGFWTLTYFVLGLALMADGVE 182
QY 135 -----VCEDDWKAPLTISLFVGVLLGSFSGQLSDRFGKRNKLVFTVMQMG 180
DB 183 VFVGVFLPSAEKDMCLSDSNKMGMLGLIVVLGMVGAFLMGGLADRLGRQCLLSLSV 242
QY 181 TGSFSLQIFSKNFEMFVFLVLMGMQISNYVAAFLVGLTEILGKSVRIIFSLG----- 234
DB 243 SVFAFFSSFGVGFTFLFCRLLSGVG-----IGGSIPIVFSFSEFLAQE 287
QY 235 -----VCIFYAFG-----YMLPLF-----AYFIQRMRLMLVALTMPGVLC 270
DB 288 KRGEHLWLCMFWMVIGVYAAAMAWAIIPHYGWSFGMSAYQFHSWRVFLVFAFSPVFA 347
QY 271 VALWVFIPESPRWLISQGRFEAEVIRKA---AKANGIVVPSTIFDPSSELDLSSKKQ 326
DB 348 IGALTTPQSPREFLENGKDEAMVILKQVHDTNMRAG--HPERVFSVTHIKTI-----H 401
QY 327 QSHNILDL-----LRTWNI-----RMVTIMSIMLWMTISVGY 358
DB 402 QEDELIEIQSDTGTWQRCVGRALSLGGVWGNFLSCFSPRYRITLMMGVWFTMSFSY 461
QY 359 FGLSLDTPNL-----FLSANVEVPAYVLAWLILLOYPRLRYSMATALELGSVL-- 414
DB 462 YGLTVWPPDMIRHLOADVAAARTKVPGERVHVTFTLENOIHRGGQYFNDKFIGLRL 521
QY 369 -----HGDIFFVNC----- 376
DB 522 KVSFEDSLFECEYFEDVTSSNTFFRNCTFINTVFYNTDLFEYKFNLSRLVNSTFLHNKE 581
QY 377 -----FLSANVEVPAYVLAWLILLOYPRLRYSMATALELGSVL-- 414
DB 582 GCPDVTGTGEGAYMYVFSFGTLAVLPNGIVSALLMDKIGRLMLA-----GSSVLSC 636
QY 415 ---LFMQLVPPDLYLATVLVMVKGFGVT--AAFSMYYVYVYTAELYPTVVRNMGVGSSTAS 470
DB 637 VSCFELSGNSESAMIA-LICLFG--GVSIASWALDVLTVELYPSDKRTTAFGLNALC 693
QY 471 RLGSIL--SPYFVLGAYDRFLPYILMGLSLTILTLTLFLPESFGTPLPOTIDOMLR 526
DB 694 KLAALVGISIFTSFVG-----ITKAAPILFASALALGSLALKLPETRGVQLQ 742
RESULT 13
YFIG_BACSU
ID YFIG_BACSU STANDARD; PRT; 482 AA.
AC P34723;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN GLVBC 3'REGION.
GN YFIG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=96262713; PubMed=8704981;
RA Yamamoto H., Uchiyama S., Fajar A.N., Ogasawara N., Sekiguchi J.;
RT "Determination of a 12 kb nucleotide sequence around the 76 degrees
RT region of the Bacillus subtilis chromosome.";
RL Microbiology 142:1417-1421(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; D50543; BAA09111.1; -;
 DR EMBL; Z99108; CAB12655.1; -;
 DR Subtilist; BG11854; yfig.
 DR InterPro; IPR001066; -;
 DR Pfam; PF00083; sugar_tr; 1.
 DR PRINTS; PS00171; SUGTRNSPRT
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
 KW Hypothetical protein; Transp; Transmembrane.
 FT TRANSMEM 30 50 POTENTIAL.
 FT TRANSMEM 60 80 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 121 141 POTENTIAL.
 FT TRANSMEM 156 176 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 264 284 POTENTIAL.
 FT TRANSMEM 302 322 POTENTIAL.
 FT TRANSMEM 332 352 POTENTIAL.
 FT TRANSMEM 353 373 POTENTIAL.
 FT TRANSMEM 401 421 POTENTIAL.
 FT TRANSMEM 424 444 POTENTIAL.
 SQ SEQUENCE 482 AA; 52756 MW; 90851C4F4C48EE01 CRC64;

Query Match 8.9%; Score: 256; DB 1; Length 482;
 Best Local Similarity 23.9%; Pred. No. 1.le-09;
 Matches 105; Conservative 83; Mismatches 167; Indels 84; Gaps 19;
 QY 144 LRTSLFFVGLGSGTSGOLSDRGKRN-----VLFV--TMGMOTGFSFLQIFSKNPEM 195
 DB 64 LVASSLLGAAFGAMFGRLSDRGKRTILYLLALLFAATLGC-----TFSPNASV 115
 QY 196 FVFLVFLVGM--GQISNYVAAVFLGTEILGKSVRI-----FSTLGVCIFVAFGVMVPL 248
 DB 116 MIAFPELLGLAVGCASVVTPTFLAETSPAERRGRIVTQNELMIVIGQLLAYTFNAIGST 175
 QY 249 FAYFIDRMLVALTMPCLVCAVW---FIPESPRMLISOGFEFAEVLIRKAAKANG 305
 DB 176 MGESANWRYMLVILTPAV---LWFGMLVPSPRMLAAKGRMDALVLRQIRE--- 229
 QY 306 IVPSTIFDPSSELQDI-----SSKKQOSHNIIDLLRTWNIRMTI---MSIMLWMT 353
 DB 230 -----DSQAQOEIKETKHAIEGTAKKAGFH---DFQEPWIRRLIFGIGIAIVQGIT 278
 QY 354 --ISVGYFGLSL-----DTNPLHGDIFVNCFLSAMVEVPAYVLANILLQLYLPFR--- 400
 DB 279 GVNSIMYGTIELREAGFQTEAALIGNI-----ANGVISVIAVIFGILLGKVRPRPMLI 333
 QY 401 ---YSNATALGGLGSLVLLFMOLVPPDLYLATVLMVGRFGVTAAFSVMVYVTAELPTV 457
 DB 334 IGOIGTWATLLIGLISVLEGTALPYVVLSTILFLAQFQTAISTVTWMLSEIFPMH 393
 QY 458 VRNMGVGVSS---TASRLGSILSPYFV-YLGAYDRFLPYILMGSTITLITLFLPES 512
 DB 394 VRGLGMSITFLWTANFLIGFTFPIILLNHIGMSATFFIFVAMNLAAIL--FVKKYVPET 451
 QY 513 FGTPLEPDTIDOMLRVKGMK 531
 DB 452 KGRSL-EQLEHSPROYGR 469
 RESULT 14
 YDJE_ECOLI
 ID YDJE_ECOLI STANDARD; PRT; 452 AA.
 AC P38055; P77244;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN ANSA-CAPA INTERGENIC
 GN REGION.
 GN YDJE.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RC SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
 RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,
 RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
 RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
 RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
 RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RC SEQUENCE OF 360-452 FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=89357501; PubMed=2670682;
 RA Jarlstrom P.G., Bezjak D.A., Jennings M.P., Beacham I.R.;
 RT "Structure and expression in Escherichia coli K-12 of the
 RT L-asparaginase I-encoding ansA gene and its flanking regions.";
 RL Gene 78:37-46(1989).
 RN [4]
 RC IDENTIFICATION.
 RP MEDLINE=95075659; PubMed=7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RT bacterial genome.";
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; AE000272; AAC74839.1; -;
 CC EMBL; D90820; BAA15560.1; -;
 CC EMBL; D90821; BAA15567.1; -;
 CC EMBL; M26934; -; NOT_ANNOTATED_CDS.
 CC Ecogene; EG12369; ydJE.
 CC InterPro; IPR001066; -;
 CC Pfam; PF00083; sugar_tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
 KW Hypothetical protein; Transp; Transmembrane; Inner membrane.
 FT TRANSMEM 23 43 POTENTIAL.
 FT TRANSMEM 60 80 POTENTIAL.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 167 187 POTENTIAL.
 FT TRANSMEM 264 284 POTENTIAL.
 FT TRANSMEM 299 319 POTENTIAL.
 FT TRANSMEM 327 347 POTENTIAL.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 14:03:35 ; Search time 44.83 Seconds
(without alignments)
1643.852 Million cell updates/sec

Title: US-09-521-195-3
 perfect score: 2883
 Sequence: 1 MRDYDEVTAFLGWGPFORL.....HTRMLKDGERPTILKSTAF 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : SPREMBL_16.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_protein.*
12: sp_unclassified.*
13: sp Vertebrate.*
14: sp_virus.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query % | | DB | ID | Description |
|------------|--------|---------|--------|----|--------|---------------------|
| | | Match | Length | | | |
| 1 | 2295.5 | 79.6 | 564 | 11 | Q9WNT6 | Q9wtnt6 mus musculu |
| 2 | 2227 | 77.2 | 551 | 4 | Q9H015 | Q9h015 homo sapien |
| 3 | 2218 | 76.9 | 551 | 4 | Q14546 | Q14546 homo sapien |
| 4 | 2157 | 74.8 | 553 | 11 | Q9R141 | Q9r141 rattus norv |
| 5 | 2148 | 74.5 | 553 | 11 | Q9Z306 | Q9z306 mus musculu |
| 6 | 885 | 30.7 | 548 | 5 | Q9VCA2 | Q9vca2 drosophila |
| 7 | 881 | 30.6 | 548 | 5 | Q01384 | Q01384 drosophila |
| 8 | 816 | 28.3 | 567 | 5 | Q9VCA3 | Q9vca3 drosophila |
| 9 | 777 | 27.0 | 568 | 5 | Q9U539 | Q9u539 caenorhabdi |
| 10 | 777 | 27.0 | 576 | 5 | Q02270 | Q02270 caenorhabdi |
| 11 | 753.5 | 26.1 | 561 | 5 | Q9VG16 | Q9vg16 drosophila |
| 12 | 740 | 25.7 | 554 | 4 | Q9NQD4 | Q9ngd4 homo sapien |
| 13 | 738.5 | 25.6 | 556 | 4 | Q75751 | Q75751 homo sapien |
| 14 | 738 | 25.6 | 554 | 4 | Q15395 | Q15395 homo sapien |
| 15 | 734 | 25.5 | 554 | 4 | Q15245 | Q15245 homo sapien |
| 16 | 733 | 25.4 | 593 | 11 | P70485 | P70485 rattus norv |
| 17 | 728 | 25.3 | 593 | 11 | P97558 | P97558 rattus norv |
| 18 | 727 | 25.2 | 555 | 4 | Q15244 | Q15244 homo sapien |
| 19 | 724 | 25.1 | 555 | 11 | Q9R0W2 | Q9r0w2 rattus norv |

ALIGNMENTS

[illegible]

QY 181 TGFSLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 240
 Db 181 TGFSLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 240
 QY 241 FGYVLPFLFAYFIRDRMMLLVALTMGVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 300
 Db 241 FGYVLPFLFAYFIRDRMMLLVALTMGVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 300
 QY 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLRTWNRMTYIMSLMWTISVG 357
 Db 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLRTWNRMTYIMSLMWTISVG 357
 QY 358 YFGLSLDTPLNHLGDIIVNCFLSAMVVPAYVLAWLILQYLPRYSMATALFLGGSVLLFM 417
 Db 361 YFGLSLDTPLNHLGDIIVNCFLSAMVVPAYVLAWLILQYLPRYSMATALFLGGSVLLFM 417
 QY 418 QLYPPDLIYLAIVLVMGVGFYTAASVYVYTAELIYPTVVRNMGVSVSTASRLGSIIS 477
 Db 421 QLYPPDLIYLAIVLVMGVGFYTAASVYVYTAELIYPTVVRNMGVSVSTASRLGSIIS 477
 QY 478 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVKGKHKRTPS 537
 Db 481 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVKGKHKRTPS 537
 RESULT 2
 Q9H015 PRELIMINARY; PRT; 551 AA.
 ID Q9H015
 AC Q9H015
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE INTEGRAL MEMBRANE TRANSPORT PROTEIN.
 GN UT2H.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Spritzberger F., Gruendemann D., Schoemig E.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y09881; CAAT1007.1;
 SQ SEQUENCE 551 AA; 62155 MW; C827A99AA78C9443 CRC64;

Query Match 77.2%; Score 2227; DB 4; Length 551;
 Best Local Similarity 76.3%; Pred. No. 3e-133;
 Matches 425; Conservative 57; Mismatches 69; Indels 6; Gaps 3;

QY 1 MRDYDEVTAFLGEGPQRLIFLLSASITPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
 Db 1 MRDYDEVTAFLGEGPQRLIFLLSASITPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
 QY 61 AWRNHTVPLRLDGRVPHSCRYRLATIANFSALEGPGRDVLGQLEQESCLDGWEFS 120
 Db 61 AWRNHTVPLRLDGRVPHSCRYRLATIANFSALEGPGRDVLGQLEQESCLDGWEFS 120
 QY 121 QDVYLSVTIVTEWNLVCEDDWAKPLTISLFFVGVLLGSGTISGQSDRFRGKNVLFVTMGQ 180
 Db 121 QDVYLSVTIVTEWNLVCEDDWAKPLTISLFFVGVLLGSGTISGQSDRFRGKNVLFVTMGQ 180
 QY 181 TGFSLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 240
 Db 181 TGFSLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 240
 QY 241 FGYVLPFLFAYFIRDRMMLLVALTMGVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 300
 Db 241 FGYVLPFLFAYFIRDRMMLLVALTMGVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 300
 QY 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLRTWNRMTYIMSLMWTISVG 357
 Db 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLRTWNRMTYIMSLMWTISVG 357
 QY 358 YFGLSLDTPLNHLGDIIVNCFLSAMVVPAYVLAWLILQYLPRYSMATALFLGGSVLLFM 417
 Db 361 YFGLSLDTPLNHLGDIIVNCFLSAMVVPAYVLAWLILQYLPRYSMATALFLGGSVLLFM 417
 QY 418 QLYPPDLIYLAIVLVMGVGFYTAASVYVYTAELIYPTVVRNMGVSVSTASRLGSIIS 477
 Db 421 QLYPPDLIYLAIVLVMGVGFYTAASVYVYTAELIYPTVVRNMGVSVSTASRLGSIIS 477
 QY 478 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVKGKHKRTPS 537
 Db 481 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVKGKHKRTPS 537

QY 361 LSLDTPNLHGDIIVNCFLSAMVVPAYVLAWLILQYLPRYSMATALFLGGSVLLFMQ 420
 Db 359 LSLDTPNLHGDIIVNCFLSAMVVPAYVLAWLILQYLPRYSMATALFLGGSVLLFMQ 418
 QY 421 PDLYYLAIVLVMGVGFYTAASVYVYTAELIYPTVVRNMGVSVSTASRLGSIIS 480
 Db 419 PDLYYLAIVLVMGVGFYTAASVYVYTAELIYPTVVRNMGVSVSTASRLGSIIS 478
 QY 481 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVKGKHKRTPS 540
 Db 479 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVKGKHKRTPS 535
 QY 541 MKDGOERPTILKSTAF 557
 Db 536 DSMETEENPKVL-ITAF 551
 RESULT 3
 O14546 PRELIMINARY; PRT; 551 AA.
 ID O14546
 AC O14546
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE POLYSPECIFIC ORAGANIC CATION TRANSPORTER.
 GN OCTN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tamai I., Yabuuchi H., Nezu J., Sai Y., Oku A., Shimane M., Tsuji A.;
 RL "Cloning and characterization of a novel human pH-dependent organic
 cation transporter, OCTN1.";
 RL FEBS Lett. 419:107-111(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
 DR EMBL: AB007448; BAA23356.1;
 DR InterPro; IPR001066;
 DR Pfam; PF00083; sugar_tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 KW Transmembrane.
 SQ SEQUENCE 551 AA; 62176 MW; F5903421C789F60A CRC64;

Query Match 76.9%; Score 2218; DB 4; Length 551;
 Best Local Similarity 75.9%; Pred. No. 1.1e-132;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

QY 1 MRDYDEVTAFLGEGPQRLIFLLSASITPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
 Db 1 MRDYDEVTAFLGEGPQRLIFLLSASITPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
 QY 61 AWRNHTVPLRLDGRVPHSCRYRLATIANFSALEGPGRDVLGQLEQESCLDGWEFS 120
 Db 61 AWRNHTVPLRLDGRVPHSCRYRLATIANFSALEGPGRDVLGQLEQESCLDGWEFS 120
 QY 121 QDVYLSVTIVTEWNLVCEDDWAKPLTISLFFVGVLLGSGTISGQSDRFRGKNVLFVTMGQ 180
 Db 121 QDVYLSVTIVTEWNLVCEDDWAKPLTISLFFVGVLLGSGTISGQSDRFRGKNVLFVTMGQ 180
 QY 181 TGFSLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 240
 Db 181 TGFSLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 240
 QY 241 FGYVLPFLFAYFIRDRMMLLVALTMGVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 300
 Db 241 FGYVLPFLFAYFIRDRMMLLVALTMGVLVLMGQISNYVAAFVLGTEILGKSVRIIFSTLGVCIYA 300
 QY 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLRTWNRMTYIMSLMWTISVG 357
 Db 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLRTWNRMTYIMSLMWTISVG 357
 QY 358 YFGLSLDTPLNHLGDIIVNCFLSAMVVPAYVLAWLILQYLPRYSMATALFLGGSVLLFM 417
 Db 361 YFGLSLDTPLNHLGDIIVNCFLSAMVVPAYVLAWLILQYLPRYSMATALFLGGSVLLFM 417
 QY 418 QLYPPDLIYLAIVLVMGVGFYTAASVYVYTAELIYPTVVRNMGVSVSTASRLGSIIS 477
 Db 421 QLYPPDLIYLAIVLVMGVGFYTAASVYVYTAELIYPTVVRNMGVSVSTASRLGSIIS 477
 QY 478 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVKGKHKRTPS 540
 Db 479 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVKGKHKRTPS 535
 QY 541 MKDGOERPTILKSTAF 557
 Db 536 DSMETEENPKVL-ITAF 551

```
QY 301 AKANGIVVPSTFDPSLQDLSSKKQSHNLDLRTNIRMTIMSLMWTISVGYFG 360
Db 301 AKNNVPAVAFD--SVEELNPKQKAFILDLPETRNAITMTIMSLMWTISVGYFA 358
QY 361 LSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLLOYLPRRYSMATALFLGSGVLLPMOLV 420
Db 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIAAVLFWGGVLLFIOLV 418
QY 421 PPDLYLATVLMVVGKFGVTAAFSMVYVTAELYPTVVRNMGVSGVSSASRLGSLSPVF 480
Db 419 PVDYFSLGLWLGKFGITSFAFSLYVTAELYPTVVRNMGVSGVSSASRLGSLSPVF 478
QY 481 VYLGAIDRELPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 540
Db 479 VYLGAIDRELPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 540
QY 541 MLKDGQERPTILKSTAF 557
Db 536 DSMETEENPKVL-ITAF 551

RESULT 4
Q9RI41 PRELIMINARY; PRT; 553 AA.
AC Q9RI41;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ORGANIC CATION TRANSPORTER OCTN1.
GN OCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu X., Wang H., Leibach F.H., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of OCTN1,
RT an organic cation transporter";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF169831; AAD46922.1;
DR InterPro; IPR001066;
DR Pfam; PF00083; sugar tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW TRANSMEMBRANE.
SQ SEQUENCE 553 AA; 62362 MW; E26C8155768A14AD CRC64;
```

```
Query Match 74.8%; Score 2157; DB 11; Length 553;
Best Local Similarity 72.7%; Pred. No. 8.1e-129;
Matches 405; Conservative 65; Mismatches 83; Indels 4; Gaps 2;

QY 1 MRDYDEVTAFLGEMGPFQRLIFELLASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
Db 1 MRDYDEVTAFLGEMGPFQRLIFELLASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
QY 61 AWRNHTVPLRLDGRVPHSCRRYRLATANFSALEPCRDVDLQLEQESCLDGEWS 120
Db 61 AWRNHTVPLRLDGRVPHSCRRYRLATANFSALEPCRDVDLQLEQESCLDGEWS 120
QY 121 QDVLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGKKNVLFVTMGQ 180
Db 121 KDVLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGKKNVLFVTMGQ 180
QY 181 TGFSEFLOIFSKNFEMFVFLVGMGQISNYVAAFVLGTEILGKSVRIITFLGVCIFYA 240
Db 181 TGFSEFLOIFSKNFEMFVFLVGMGQISNYVAAFVLGTEILGKSVRIITFLGVCIFYA 240
QY 241 FGYWVLPFAFIRDMRMALLVATMPGVLCVALLMFWIPESPRWLLISQGRFEAEVIRKA 300
Db 241 IGYWVLPFAFIRDMRMALLVATMPGVLCVALLMFWIPESPRWLLISQGRFEAEVIRKA 300
QY 301 AKANGIVVPSTFDPSLQDLSSKKQSHNLDLRTNIRMTIMSLMWTISVGYFG 360
Db 301 AKNNVPAVAFD--SVEELNPKQKAFILDLPETRNAITMTIMSLMWTISVGYFA 358
```

```
Db 301 AKNGIMAPAVIEDPDLQELNSLKQKVFILDLFKTRNIATITVMSVLMWLTISVGYFA 360
QY 361 LSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLLOYLPRRYSMATALFLGSGVLLPMOLV 420
Db 361 LSLNPNLHGDIIVNCFLSGIEVPAYFTAWLLRTLPRIIACVLFWGGVLLVQVV 420
QY 421 PPDLYLATVLMVVGKFGVTAAFSMVYVTAELYPTVVRNMGVSGVSSASRLGSLSPVF 480
Db 421 PEDYFVSLGLWLGKFGITSFAFSLYVTAELYPTVVRNMGVSGVSSASRLGSLSPVF 480
QY 481 VYLGAIDRELPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 540
Db 481 VYLGAIDRELPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 540
QY 541 MLKDGQERPTILKSTAF 557
Db 538 VSMDEENPKVL-ITAF 553

RESULT 5
Q9Z306 PRELIMINARY; PRT; 553 AA.
AC Q9Z306;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ORGANIC CATION TRANSPORTER.
GN OCTN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Nezu J.;
RT "Mouse OCTN1: Polyspecific organic cation transporter";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AB016257; BAA36626.1;
DR InterPro; IPR001066;
DR Pfam; PF00083; sugar tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW TRANSMEMBRANE.
SQ SEQUENCE 553 AA; 62290 MW; C4D66BC061398653 CRC64;
```

```
Query Match 74.5%; Score 2148; DB 11; Length 553;
Best Local Similarity 72.2%; Pred. No. 3e-128;
Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;

QY 1 MRDYDEVTAFLGEMGPFQRLIFELLASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
Db 1 MRDYDEVTAFLGEMGPFQRLIFELLASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
QY 61 AWRNHTVPLRLDGRVPHSCRRYRLATANFSALEPCRDVDLQLEQESCLDGEWS 120
Db 61 AWRNHTVPLRLDGRVPHSCRRYRLATANFSALEPCRDVDLQLEQESCLDGEWS 120
QY 121 QDVLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGKKNVLFVTMGQ 180
Db 121 KDVLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGKKNVLFVTMGQ 180
QY 181 TGFSEFLOIFSKNFEMFVFLVGMGQISNYVAAFVLGTEILGKSVRIITFLGVCIFYA 240
Db 181 TGFSEFLOIFSKNFEMFVFLVGMGQISNYVAAFVLGTEILGKSVRIITFLGVCIFYA 240
QY 241 FGYWVLPFAFIRDMRMALLVATMPGVLCVALLMFWIPESPRWLLISQGRFEAEVIRKA 300
Db 241 IGYWVLPFAFIRDMRMALLVATMPGVLCVALLMFWIPESPRWLLISQGRFEAEVIRKA 300
QY 301 AKANGIVVPSTFDPSLQDLSSKKQSHNLDLRTNIRMTIMSLMWTISVGYFG 360
Db 301 AKNNVPAVAFD--SVEELNPKQKAFILDLPETRNAITMTIMSLMWTISVGYFA 358
```

```

Db 301 AKNSIVAPAGIDFPLEQLQSLKQKVILDLFRNIAITTVMAVLMWLMTSVGYFA 360
QY 361 LSLDTPLNLDGDFVNCFLSAMVEVPAYVLAWLILLOVLPYRYSNATFALFGGSSVFLPMQVLY 420
Db 361 LSNVNPVNLHGDVYVNCFLSLIEVPAYFTAWLLRLTPRYIAGVLFWGGVLLIQVY 420
QY 421 PPDLXYLATVLMVGVKFGVTAAPSMVYVYTAELPTVVRNMGVSVSTASRLGSLSPYF 480
Db 421 PEDYNEVSGVLMLGRFGITSAFSLMYVYTAELPTVLRNMAVGITSMASRVGSIIAPYF 480
QY 481 VYLGAVDRFLPYLMLGSLTILTAFLTLFLPESFGTLPDITDMLRVKGMKHKTPSHTR 540
Db 481 VYLGAVNRLPYLMLGSLVLIIGITLFFPESFGVTLPENLEQMKVGRFCGK---KST 537
QY 541 MLKDGOERPTILKSTAF 557
Db 538 VSDREESPKVL-ITAF 553

RESULT 6
Q9VCA2 PRELIMINARY; PRT; 548 AA.
AC Q9VCA2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE ORCT PROTEIN.
GN ORCT OR CG6331.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej K.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Deichler A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AE003747; AAF56271.1; -.
DR FlyBase; FBgn0019952; Orct.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
DR Transmembrane.
SQ SEQUENCE 548 AA; 61002 MW; 08D7F97599B477AF CRC64;

Query Match 30.7%; Score 885; DB 5; Length 548;
Best Local Similarity 35.1%; Pred. No. 2.le-48;
Matches 198; Conservative 106; Mismatches 206; Indels 54; Gaps 9;

QY 4 YDEVTAFLGEGPQRLTFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVP----DAANLS 59
Db 3 YDDVTIHLGEGPYQKRIYYLLCLPAIVCAFHKLAGVLLAKPDPFCALPYENGSIYELS 62
QY 60 SAWRNHVPRLRDGVRPHSCRYRLATTANFSALEGRDVL-----GQL----- 108
Db 63 PHLNLSYPENER-----CSYY-----DVDYTEYLNGSIIPRSSN 97
QY 109 QEESCLDGWEFSQDVLSTIVTEWNLVCEDDWKAPLTISLFFVGLGSGISGOLSDRFQ 168
Db 98 ETKTC-SSYVYDRSKYLSAVTEWNLVCSRLSLSATSDSLFLGLVGLSLIFGQMSDKLG 156
QY 169 RKNVLFVTMGMTQSFSLQIFSKNFEMFVFLVFLVGMGQISNYVAAPVLGTEILGKSVRI 228
Db 157 RKPTFFASLVLLQIFGLVLAAPVPEYFTYISRTMVGATTSGVFLVAVVALEVMVGSYR- 215
QY 229 IFSTLGVCIYAFGYMVLPLFAYFIRDRMLLVALLTPGVLVGLVALWVFIPESPRLISQ 288
Db 216 LFAGVAMQMFSGVFMLTAGFAYFIHWRWLQITAILPGLLFLCYWITIPESARWLMLKG 275
QY 289 RFEEAEVIRKAANGIVVPTIFD-----PSEQLDLSKKQKQSHNTLDLLRTWNIR 341
Db 276 RKDEAFVIEKAENKENVPEIYELVDVEAEKKQDEMAASQAPANTVFDLLRYPNLR 335
QY 342 MVTMTSLMWNITISGVYGLSLDTPNLHGDFIVNCFISAMVEVPAYVLAWLILQYLPYRY 401
Db 336 KRTLLIFDFNVSGVYGLSWNTNLLGNLQNFMSGAVEIPGYTLLLTLLNRWGRS 395
QY 402 SMATALFLGSLVLLFMOLVPPDLYLATVLMVCKFGVTAAPSMVYVYTAELPTVVRNM 461
Db 396 ILCTMMVAGLSLTLATTFVPSDMNWLIVACAMIGKLAITSYGTIYFSAEQPTVVRNV 455
QY 462 GVGVSSTASRLGSLSPYFVFLGAYDRFLPYLMLGSLTILTAFLTLFLPESFGTLPDIT 521
Db 456 GLGASSMVARVGGILAPYLLGLGEINRPLLIICGALSITAGLSLLPPTLKNPPEITI 515
QY 522 DQMLRVKGMKHKTPSPHTRMLKDG 545
Db 516 ED-----GENFGKKPAQETAEEG 534

RESULT 7
ID 001384 PRELIMINARY; PRT; 548 AA.
AC 001384;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE ORGANIC CATION TRANSPORTER.
GN ORCT OR CG6331.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]

```


QY 62 WRNHTVPLRLDGRVPHSCRYRLATIANFSALGLEPGRDVLQLEQESCLDQWERSQ 121
Db 76 LTNDTQIL-----SKOYNETQNVFRAFTSAP-VDTYSRISLVPQNGWDYDN 124
QY 122 DVLSTIVTEWNLVCDWKAPLTSLFFVGVLLGSFISGSLDRFGKKNLFLVTMGHQT 181
Db 125 STYLSLVTETNVLCDQQAWEISTSPVGSFIGNCLFGYVADKFGRRRFFVILTVLI 184
QY 182 GFSLQIFSKNFMFVFLVFLVGMQISNYVAAFVLGTGILGKSVRIIFSLGVCIFYAF 241
Db 185 VCGTASSPAKDIESTFILLRFFGTFLAPALFQIPFTICMEFGMNGSR-IFSGILMTSLFFGA 243
QY 242 GYVLPFLFAFYFTRWMLLVALTMGVLCVALWVLPSPRWLSISQGFEEAEVLIIRAA 301
Db 244 AMALLGVWAMFRRRQLTFFCNPAFAFYIYFFLPSPRWSSVSGWADAKKOLKIA 303
QY 302 KANG-----IIVPSTIFDPSLOLSSKKQ--QSHNILLRLTWNIRMTIMSLMWTISV 356
Db 304 KNGKSNVDVDELVDMSKMHQNAEKEKTRSHNVTDLFKTPNLRKTLIVTYIWMVNAI 363
QY 357 GYFGLSLDPNHLGDIFFVNCFLSVMVEPVAVVLAMLLQYLPRRYSMATATFLGGSVILF 416
Db 364 IYNGLTNLVSNLPLVDYWSFIINGAVELPGYFVWVPLLCQAGRRWTLAATMIVCGIGCVS 423
QY 417 MOLVPPDLXYLATLVLMVCKFGTAFAFMSVYVYTAELPTVVRNMGVGVSSPASLGSIL 476
Db 424 AMFMDGYPWLVASFICGKFGVGSFAVIYIFAGELPTVVRNMGVGVSSPASLGSIL 483
QY 477 SPYFVYGLDYAPFLYILMGSTITLITAILTLPLPSFGFTPLDPTD 522
Db 484 APHIVNLGKTVKTLPLINGLMAISAGILTFELPTELGAFLPMTIE 529
RESULT 11
QYV6L6
ID Q9V6L6 PRELIMINARY; PRT; 561 AA.
AC Q9V6L6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CG4630 PROTEIN.
GN CG4630.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclele J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL; AE003819; AAF58407.1; -.
DR FlyBase; FBgn0033809; CG4630.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar.Lr; 1.
DR PF00083; sugar.Lr; 1.
SQ SEQUENCE 561 AA; 62041 MW; 573643481595788F CRC64;
Query Match 26.1%; Score 753.5; DB 5; Length 561;
Best Local Similarity 33.2%; Pred. No. 4.4e-40;
Matches 184; Conservative 120; Mismatches 200; Indels 51; Gaps 14;
QY 3 DYDEVTAFLGWSGFORLIFLFLLSASIIPIGFTGLSSVFLIATPEHRCRVPDANLSSA- 61
Db 14 DFDLILVEIGFGRFORNYLLICLPVLFAAANSLSVFTAGSPTRYCYRYPEDKLVDAE 73
QY 62 ---WRNHTVP-LRLDGRVPHSCRYRLATIANFSAL--GLEPGRDVLQ----- 107
Db 74 YGANVWSIAPGWSKRGHPTSTCERF-----VANGHLESSSDPWSAWPLDQCAENFT 129
QY 108 LEQESCLDQWERSQDQVYLS---TIVTEWNLVLC-EDDMKAPLITISLFFVGVLLGSFISQGL 163
Db 130 TETERC-----NQFVYGSSTIVQQWGLQCPENLKLAFVGTIFHAGLVGVGTALSGYL 183
QY 164 SDRGRKNV-LFVTMGMO-TGFSFLQIFSKNFMFVFLVFLVGMQISNYVAAFVLGTETI 221
Db 184 ADYGRKHIFLFCIVFMALTGVA--QALSWDYISLFFALLNAVGTSGVYPLAFIIGVM 241
QY 222 LGKSVRIIFSTLGVCIYFAGYVMVLPFLFAYFIRDRWMLLVALTMPGVLCVALWVFIESP 281
Db 242 VGPKRREM-SSIVUNYFVAGGALLGL--SVFLPDWRQLQALSVPLPICVAYFWLWVPSV 299
QY 282 RWLISQGRFEEAEVLIIRKAANKANGIVVPSTIFDPSLOLSSKKQSHN----- 330
Db 300 RWLLARNRREGAGVLIIRAAKNRRDISVELMASFKQQLDAETGQEDDVEGGLHVRKDD 359
QY 331 -----ILDLLRTNIRMTIMSLMWTISVGYTGLSLDTNHLGDIFFVNCFLSVMVEVP 385
Db 360 KIWLAVKEVARSHILMGRYAILLIWNAIVYGLSNATSLGKNYINLALVCLVEIP 419
QY 386 AVYLAWLLOLVLPRYSMATATFLGGSVLLFMQLVPLDLYLATVLMVGVKFGVTAAFSM 445
Db 420 GYSLAWLFLRFRGRVALUGSLLSITCIVAGSFTVIGANWLVTLFLVGLKITSSFAV 479
QY 446 VVYVYTAELPTVVRNMGVGVSSPASLGSILSPYFVYGLDYAPFLYILMGSTITLITAIL 505
Db 480 IYFTTAEMPTVIRSGGVGMSTPFAFGAMLAPFVPLLSAYYDPLPLLLFGLTSLVAGLL 539
QY 506 TILFSPESFGTPLDPT 520
Db 540 SLLLPETFNRLKLPDT 554
RESULT 12
ID Q9N0D4 PRELIMINARY; PRT; 554 AA.
AC Q9N0D4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)


```
FT CARBOHYD 322 322 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 556 AA; 61279 MW; C3CA2D7DD21C658 CRC64;

Query Match 25.6%; Score 738.5; DB 4; Length 556;
Best Local Similarity 33.0%; Pred. No. 3.9e-39;
Matches 189; Conservative 117; Mismatches 214; Indels 53; Gaps 18;

QY 1 MRDYDEVTAFLGEGWGFQRLIFELLSASIIIPNGFTGLSSVFLIATPEHR-CRVPDAANLS 59
DB 1 MPFDEALQVGEFGFRFRRVLLCLTGVTAFILFVGVGTOTPDHYWCRGPSAAALA 60
QY 60 S-AWR-----NHTVPLRLRDGVEPH-----SCRRVRLATIANFSA-----LGLE 98
DB 61 ERGWSPEEWNRTAP--ASRGPEPERGRGORY--LLEAANDSASATLSACDPLAAF 117
QY 99 PGDRVDLQLEQBSCLDGNFESQDVYLSITVTENWLVCEDDWKAPLITSLFFVGVLLGSF 158
DB 118 PNRSAPL-----VPCRGGWRYAQ--AHSTIVSEFDLVCVNAWMLDLTQAILNLGLTGA 170
QY 159 ISQLSDRGKRVLVMTGMQTFSELOI-PSKNFEMFVFLVVLGVMQISNYVAAYVL 217
DB 171 TLGYAADRYGR-ITVIYLLSCLGVGTGVVVAFAFPVFIREFLQGVFGKGTWMTCYVI 229
QY 218 GTILGKSVRIEFTSLGVCII--FYAFGYMVLPLFAFYFIRDRWMLLVALTMPGVLCVALW 275
DB 230 VTEIVSKQRI---VGIIVQMEFTLGIILPGIAFYFNWOGIQAIALPSEFLYYW 286
QY 276 FIPESRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSELODLSKKKQSHNILDLL 335
DB 287 VVPSRWLITRKKGKALKILRIAKCKNGKYLSSNY---SEI-TVTDEEVNPSFLDIV 342
QY 336 RTNWRVMTIMSLMWTISVGVFGLSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLLQ 395
DB 343 RTPQMRKCTILMFAMFTSAVYVQGLVMRLGIIGNLYIDFISGVLPGLALLITTE 402
QY 396 YLPRLRYSMATLFGSGVLLFMQLVPPDLYLATVLMVGVKFGVTAFAFSVYVYTAELXP 455
DB 403 RLGRLLPFAASNIVAGVACLVTAFLPEGIANLRTVATLGRGITWAFIIVLVNSELXP 462
QY 456 TVVRNMGVSVSTASRLSGTSLSPYFY-IGAYDRFLPYILMGSLTILTALTILFLPESG 514
DB 463 TTLNFGVSCSLCDFGGLIAPFLFRLAAVWLEPLIIFGLASIGGLVMLLPETKG 522
QY 515 TPLPDTIDQLRV-----KGMKHKRTP---SH 538
DB 523 IALPETVDDVEKLGSHPSCCKGRNKKTPVSRSH 555

RESULT 14
O15395 PRELIMINARY; PRT; 554 AA.
AC O15395;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE ORGANIC CATION TRANSPORTER 1.
GN HOCT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97330822; PubMed=9187257;
RA Zhang L., Dresser M.J., Gray A.T., Yost S.C., Terashita S.,
RA Giacomini K.M.;
RT "Cloning and functional expression of a human liver organic cation
RT transporter.";
RL Mol. Pharmacol. 51:913-921(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; U77086; AAB67703.1; -.
```

```
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW Transmembrane.
SQ SEQUENCE 554 AA; 61153 MW; 55206B897DE32202 CRC64;

Query Match 25.6%; Score 738; DB 4; Length 554;
Best Local Similarity 35.9%; Pred. No. 4.2e-39;
Matches 197; Conservative 86; Mismatches 215; Indels 50; Gaps 12;

QY 1 MRDYDEVTAFLGEGWGFQRLIFELLSASIIIPNGFTGLSSVFLIATPEHRCKRVPDAAN 57
DB 1 MPTVDDILQEVGSGWFKQAFILCLLSAFAF---ICVGIPLFGTTPDHHCOSPGVAE 57
QY 58 LSS-AWR-----NHTVPLRLRDGVEPHSCRRYR-----LATIA-NFSAIG 96
DB 58 LSQRCGSPAEELNATVPGGLPAGEAFLGOCRRYEVDMNQSALSACVDPLASLATNRSLHP 117
QY 97 LEPRGVDLQLEQBSCLDGNFESQDVYLSITVTENWLVCEDDWKAPLITSLFFVGVLLG 156
DB 118 LGP-----CODGWY--DTPGSSIVTEFNLCVADSKLDFQSLNAGELFG 162
QY 157 SFISQSLSDRGKRVLVMTGMQTFSELOI FSKNFEMFVFLVVLGVMQISNYVAAYV 216
DB 163 SLGVGYFADRGKRLCLLGTVLVNAVSGVLMFSPNYMSMLLFRLLQGLSVSKGNMAGYT 222
QY 217 LGTEILGKSVRIEFTSLGVCIFYAF--GYMVLPLFAFYFIRDRWMLLVALTMPGVLCVALW 274
DB 223 LITFVFGSGSR---RTVAIMYQMAFTVGLVALTGLAYALPHRWLQLAVALSLTFLFLYY 279
QY 275 WFIPESRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSELODLSKKKQSHNILD 334
DB 280 WCVPESRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSELODLSKKKQSHNILD 337
QY 335 LRTNWRVMTIMSLMWTISVGVFGLSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLL 394
DB 338 FRTPLRKRTFLMWFTDSVLYQGLILHMGATSGNLYLDFLYSALVEIPGAFIALITI 397
QY 395 QYLPRLRYSMATLFGSGVLLFMQLVPPDLYLATVLMVGVKFGVTAFAFSVYVYTAEL 454
DB 398 DRVGRYIPWAMNSNLAGAACLVMIFISPDHLWLIIMCVGRMGITTAIOMICLVNAEL 457
QY 455 PTVRNMGVSVSTASRLSGTSLSPYFY-IGAYDRFLPYILMGSLTILTALTILFLPESF 513
DB 458 PTFVRNLGVMVCSLDCIGIITFIVPRLVQWALPILFAVLGLLAAGVTLLPETK 517
QY 514 GTPLPDTI 521
DB 518 GVALPETM 525

RESULT 15
O15245 PRELIMINARY; PRT; 554 AA.
AC O15245;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE ORGANIC CATION TRANSPORTER.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RX MEDLINE=97405886; PubMed=9260930;
RA Gorboulev V.G., Ulzheimer J.C., Akhondova A.;
RT "Cloning and characterization of two human polyspecific organic cation
RT transporters.";
RL DNA Cell Biol. 16:871-881(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
```

```
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; X98332; CAA6977.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar tr. 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW Ion transport; Transmembrane.
SQ SEQUENCE 554 AA; 61187 MW; B16E200852CC4000 CRC64;

Query Match      25.5%; Score 734; DB 4; Length 554;
Best Local Similarity 35.8%; Pred. No. 7.5e-39;
Matches 196; Conservative 86; Mismatches 216; Indels 50; Gaps 12;

QY 1 MRDYDEVTAFLEGMGPQR---LIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAAAN 57
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 MPTVDDILEQVSGSGWFKQAFLILCLLSAFAF---ICVGIVFLGFTPDHHCQSPGVAE 57

QY 58 LSS--AWR-----NHTVPLRLRGREVPHSCRRYR-----LATIA-NFSALG 96
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 58 LSORCGWSPAEELNYTPVGLGPAGEAFLGQCRRYEYDWNQSAALSCVDPLASLATNRSLP 117

QY 97 LEPRGDVLAQLQECSCLDGWEFSQDVYLSITVTENNLVCEDDQKAPLTISLFFVGVLLG 156
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 118 LGP-----CQDGVY--DTPGSSIVTEFNLCADSWKLDLFQSCINAGFFEG 162

QY 157 SFISGQLSDRGRKNVLFVTMGMTGFSFLQIFSKNFEMFVVLVFLVGMGQISNYAAEV 216
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 163 SLGVGYFADRFGRKLCILGTGLVNAVSGVLMAPSPNYSMLLFRLLQGLVSKGNMAGYT 222

QY 217 LGTEILGKSVRIIFSPLGVCIFYAF--GYMVLPLFAYFTRDRMLLVALTMPGVLCAVM 274
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 223 LITEFVGSGR---RIVAIMYQMAFTVGLVALTGLAYALPHRWLQQLAVSLPTFLFLYY 279

QY 275 WFIPESPRWLISQRFEEAEVIRKAAKANGIVWPSTIFDPSELODLSKKQOOSHNIIDL 334
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 280 WCVPESPRWLLSQKRNTEAIKIMDHIAQKNGKLPADLKLMSLEEDVTEK--LSPSFADL 337

QY 335 LRTWNIRWVTIMSLMWTISVGYFGLSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLL 394
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 338 FRTPRLRKRTFIILMYLFTDSVLYQGLILHMGATSGNLYLDFLYSALVEIPGAFIALITI 397

QY 395 QYLPRRYSNATALFLGGSVLLPMOLVPPDLYLYLATVLMVWKGFGVTAAPSMVVVYTAELY 454
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 398 DRVGRIYPMAAMNLLAGACLWIFISPDHLWLNIIIMCVGRMGITITATQMICLVNAELY 457

QY 455 PTVRNMGVGSVTSARGLSILSPYEVY-LGAYDRFLPYILMGSLTILTAITLFLPESF 513
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 458 PTEVRNLGVNMCSSLCDIGGIITPFTFVRLREVWQALPLILFAVLGLLAAGVYLLLPETK 517

QY 514 GTPLPDTI 521
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 518 GVALPETM 525
```

Search completed: August 16, 2001, 14:03:36
Job time: 308 sec